UNIVERSITA' DEGLI STUDI DI PADOVA

FACOLTA' DI STATISTICA

CORSO DI LAUREA IN STATISTICA E FINANZA

TESI DI LAUREA

LA PERFORMANCE DELLE PROCEDURE DI DISAGGREGAZIONE TEMPORALE MEDIANTE INDICATORI DI RIFERIMENTO. UN'ANALISI SU 94 SERIE OECD

Relatore: PROF. TOMMASO DI FONZO

Laureando: Federico Previtera

Matricola: 462837 / SE

ANNO ACCADEMICO 2004-2005

INDICE

1- INTRODUZIONE

In economia, specialmente per le decisioni di politica economica, è spesso indispensabile avere a disposizione dati a cadenza temporale elevata e di qualità, fatto che porta alla necessità di sviluppare delle procedure che intervengano nel caso in cui i dati a disposizione abbiano frequenza bassa. Come vedremo le tecniche utili a questo fine richiedono una copiosa attività di implementazione informatica. Inoltre, è cagionevole adoperarsi affinché i metodi in questione non solo possiedano solide basi teoriche, ma siano anche facilmente utilizzabili dagli utenti interessati.

Lo sviluppo dei programmi che consentano l'utilizzo di tali tecniche è molto complesso, di conseguenza l'aggiornamento della pratica corrente a favore di migliori soluzioni può richiedere tempi anche molto lunghi.

L'oggetto di questo lavoro, la stima indiretta di serie storiche economiche trimestrali mediante modelli di regressione lineare, si inserisce in un contesto in continua evoluzione, spesso caratterizzato da un discreto *lag* temporale fra sviluppo teorico e realizzazione pratica.

Sebbene in certi casi sia decisamente preferibile avere dei conti a cadenza trimestrale anziché annuale (e, allo stesso modo, mensile anziché trimestrale), tuttavia i dati sono spesso disponibili solo con cadenza bassa perché difficilmente reperibili altrimenti e, inoltre, non è generalmente possibile avere una stima con metodi diretti: ne consegue la necessità di ricorrere ai processi di stima indiretta.

In questo lavoro tratteremo modelli di regressione atti a stimare dati a frequenza alta a partire dai disponibili, in parte temporalmente aggregati e in parte no, tramite una relazione che sfrutta appositi indicatori di riferimento; studieremo l'effettivo rendimento di venti modelli (che in seguito ridurremo a sedici),uno dei quali tuttora utilizzato dall' ISTAT e da numerosi enti statistici ufficiali, sfruttando le opzioni di stima offerte da un programma ancora in fase di sviluppo e verifica. Uno dei principali obiettivi del lavoro è la valutazione della procedura di stima ufficialmente adottata dall'Istat, sia per apprezzarne le peculiarità e gli aspetti positivi, sia per avere una prima idea dei margini di miglioramento che, con fatica relativamente ridotta, si potrebbero ottenere usando procedure di stima alternative.

Le analisi verranno condotte sulle serie storiche di un database già oggetto di studio di Moauro e Savio (2004), a cui faremo talvolta riferimento come termine di paragone dei risultati ottenuti.

Il lavoro è organizzato nel modo seguente: nel secondo capitolo tratteremo il problema della disaggregazione temporale per introdurre i metodi a cui faremo riferimento, ossia quelli in grado di fornire stime ottimali, condizionatamente a determinate ipotesi di riferimento.

Nel terzo capitolo presenteremo i dati a nostra disposizione e le metodologie di lavoro che saranno utilizzate, soffermandoci su alcuni problemi riscontrati durante la fase di elaborazione.

Il quarto capitolo fornirà una presentazione dell'analisi empirica che verrà svolta in questo progetto.

Il quinto e sesto capitolo proporranno una dettagliata analisi dei risultati ottenuti, mentre nel settimo e ultimo capitolo valuteremo singolarmente e complessivamente le performance dei vari modelli.

2- LA DISAGGREGAZIONE TEMPORALE

2.1- I METODI DI STIMA INDIRETTI

Nell'introduzione abbiamo accennato alla necessità di utilizzare procedure di stima indiretta; queste possono convenzionalmente essere distinte in due filoni fondamentali: quelle che non fanno uso di indicatori di riferimento e quelle che invece ne fanno uso.

Nel primo caso si ripartiscono i dati annuali tramite procedure matematiche partendo unicamente da questi valori noti, facendo in modo che i risultati rispettino il vincolo di aggregazione annuale. Si tratta di un approccio alquanto discutibile in quanto privo di fondamenti logici solidi, ed il suo utilizzo è consigliato quando non ci sono informazioni aggiuntive circa la serie, mentre appare ovvio che, ogni qualvolta queste informazioni ci siano, è vivamente consigliabile utilizzarle.

Se si dispone di uno o più indicatori di riferimento, ossia di una serie storica il cui andamento si ritiene che approssimi quello della serie da disaggregare, è consigliabile utilizzare procedure appartenenti al secondo dei due filoni precedentemente descritti: in questo caso l'approccio standard consiste nell'esplicitare tramite un modello di regressione lineare il rapporto fra la serie da disaggregare e gli indicatori di riferimento. Ovviamente quanto più gli indicatori di riferimento riflettono il reale andamento della serie storica oggetto di studio tanto più il processo di stima sarà preciso.

I risultati sono la conseguenza di una fase di stima preliminare dei dati e una successiva fase di aggiustamento, che distribuisce le discrepanze fra i valori trimestrali aggregati e il dato annuale noto, di modo che sia sempre rispettato il vincolo di aggregazione annuale. Queste due fasi sono integrate in un processo di stima ad un passo nei metodi ottimali (tutti quelli che prenderemo in considerazione), mentre sono distinte nei metodi non ottimali.

La fase di aggiustamento si caratterizza per poter essere sviluppata secondo numerose varianti, e tuttora non esiste un metodo di trimestralizzazione universalmente riconosciuto come più efficace.

Soffermando rapidamente l'attenzione sulle procedure di aggiustamento appartenenti al primo filone, nel corso di un anno la discrepanza tra il noto valore annuale e la somma dei quattro valori trimestrali preliminari può essere ripartita :

a- equamente tra i trimestri

b- in maniera progressiva crescente dal primo all'ultimo trimestre

c- in maniera proporzionale all'entità delle stime preliminari

Senza entrare in ulteriori dettagli (per essi, si veda Panciotti, 1965, Di Fonzo, 1987) si può sinteticamente dire che il processo di trimestralizzazione in due passi è inefficiente rispetto al processo di stima ottimale, ossia di stima in un passo, essendo caratterizzato da una maggiore varianza delle stime. Va detto che anche il previsore ottimale è la somma di un vettore di stime preliminari e di uno di aggiustamenti, ma in questo caso, la ripartizione delle discrepanze non avviene in modo arbitrario, bensì discende logicamente dalle premesse su cui il metodo di trimestralizzazione si fonda.

Noi ci muoveremo all'interno di quest' ultima categoria di modelli, tratteremo cioè stime ottimali mediante indicatori di riferimento.

2.2- LA STIMA OTTIMALE

Dato un vettore (n,1) y_0 di osservazioni annuali per un aggregato di flusso e data la matrice (4n,p) X delle corrispondenti osservazioni trimestrali per p(>1) indicatori di riferimento, vogliamo stimare gli ignoti valori trimestrali, contenuti nel vettore (4n,1) y, dell'aggregato d'interesse.

Ipotizziamo quindi che queste osservazioni trimestrali ignote siano legate agli indicatori di riferimento secondo il modello di regressione multipla

$$y = X\beta + u \tag{1}$$

in cui β è un vettore (p,1) di parametri ed u è un vettore casuale (4n,1) con E(u X) = 0 e

E(uu X) = V , assunta nota.

Si indichi con B la matrice che trasforma le 4n osservazioni trimestrali in n osservazioni annuali, premoltiplichiamo il modello (1) per B e otteniamo il modello di regressione annuale

$$y_0 = X_0 \beta + u_0 \tag{2}$$

Dove X_0 = B X è la matrice (n,p) delle somme annuali per gli indicatori di riferimento e u_0 è un vettore casuale (n,1) di media nulla e matrice di covarianza $B\ V\ B^T = V_0$

Il miglior previsore lineare in y_0 dell'ignoto vettore y, è dato da (Chow e Lin, 1971)

$$y = X\hat{\beta} + L\hat{u}_0 \tag{3}$$

dove

$$\hat{\beta} = (X_0^T\ V_0^{-1} X_0)^{-1} X_0^T\ V_0^{-1}\ y_0 \tag{4}$$

$$L = V\ B^T V_0^{-1} \tag{5}$$

$$\hat{u}_0 = y_0 - X_0\ \hat{\beta} = (\ I_n - X_0\ (X_0^T\ V_0^{-1} X_0)^{-1} X_0^T\ V_0^{-1}\ )\ y_0 \tag{6}$$

Si osserva che:

1- $\hat{\beta}$ è lo stimatore dei minimi quadrati generalizzati (GLS= *Generalized Least Squared)*

2- i valori trimestrali sono la somma di due addendi ben distinti che equivalgono indicativamente ad una fase di stima preliminare ($X\hat{\beta}$) ed uno di aggiustamento

3- il vincolo di aggregazione annuale è rispettato, come si può verificare premoltiplicando la (3) per la matrice di aggregazione temporale B.

3- PRESENTAZIONE DEI MODELLI

3.1- CHOW E LIN - AR(1)

Il modello sviluppato da Chow e Lin (1971) parte da un' ipotesi comune a tutti i modelli che vedremo anche nel seguito: l'ipotesi in questione è che gli ignoti valori di interesse e gli indicatori di riferimento (nei casi da noi considerati abbiamo sempre a disposizione un solo indicatore) siano legati da una relazione lineare del tipo:

$$y_t = \alpha + \beta x_t + u_t \quad , \quad t=1,\ldots,4n \tag{7}$$

Ciò che varia nei diversi approcci è l'atteggiamento nei riguardi della parte non deterministica: Chow e Lin infatti propongono di considerare il modello (7) caratterizzato dalla presenza di un disturbo stocastico AR(1):

$$u_t = \rho u_{t-1} + \varepsilon_t \quad , \quad t=1,\ldots,4n \quad , |\rho|<1 \tag{8}$$

Il metodo sviluppato da Chow e Lin è tuttora utilizzato dall'ISTAT e da numerosi enti statistici per i processi di disaggregazione temporale, nonostante esso mostri talvolta segni di inadeguatezza (che in seguito analizzeremo).

3.2- FERNÀNDEZ – RANDOM WALK

Fernàndez cerca di approfondire i legami esistenti tra il metodo di Denton, un metodo di aggiustamento che ha il pregio di agire in maniera semplice ed efficace, ma ha il difetto di essere un procedimento meccanico, e la soluzione ottimale di Chow e Lin; così facendo dimostra come un'appropriata scelta di V dia luogo ad un previsore ottimale di semplice elaborazione e di cui quello di Denton non è che un caso particolare.

Le premesse su cui si basa il modello di Fernàndez, ossia la presenza di un disturbo *random walk* nella (7), possono essere appropriate in diversi casi:"per serie storiche economiche quali la velocità di circolazione della moneta, i tassi d'interesse, i prezzi e la produzione industriale. In questo caso, un'ipotesi di residui random walk per alcune delle serie citate non dovrebbe essere considerata come irrealistica" .(Fernàndez, 1981, p.475)

Il modello di regressione (7) sarà quindi caratterizzato da un disturbo stocastico del tipo

$$u_t = u_{t-1} + \varepsilon_t \,, \quad t=1,...,4n \,, \; u_0 = 0 \qquad (9)$$

In fase di risoluzione il sistema avrà il pregio di avere una matrice delle covarianze degli errori nota (a meno di σ_ε^2), cosa che rende ancora più semplice un'elaborazione già alleggerita dalla non eccessiva complessità del modello.

Una differenza sostanziale rispetto al modello di Chow e Lin è data dal fatto che si presume una non stazionarietà degli errori, cosa che può risultare plausibile in molti casi; è peraltro immediato notare che il modello (3) stabilisce una relazione con disturbi white noise tra le differenze prime: operando infatti la differenza prima sulla (9) si ottiene:

$$\Delta y_t = \beta \, \Delta x_t + \varepsilon_t, \quad , \; t=2,...,4n \qquad (10)$$

Con ε_t *white noise.*

3.3- LITTERMAN – ARIMA (1,1,0)

Il modello di Litterman costituisce un ampliamento di quello, visto in precedenza, di Fernàndez.

Si tratta infatti di un modello caratterizzato da un disturbo *ARIMA(1,1,0)*, che ovviamente comprende come caso particolare il modello *Random Walk* visto in precedenza; Litterman ipotizza infatti:

$$u_t = u_{t-1} + v_t \tag{11}$$

$$v_t = \rho\, v_{t-1} + \varepsilon_t \quad , \; t=1,\ldots,4n \qquad , \quad |\rho|<1 \;,\; v_{-1} = v_0 = 0 \tag{12}$$

Il modello è più complesso di quelli visti in precedenza e richiede tempi di elaborazione informatica più lunghi, ancorché non proibitivi.

Come già visto in precedenza per Fernàndez, anche in questo caso si può dimostrare che il modello agisce sui saggi di variazione, assumendo però un disturbo stocastico di tipo AR(1) anziché *white noise*.

La stima dei parametri avviene, in ipotesi di normalità dei disturbi, secondo il metodo della massima verosimiglianza(*), massimizzando la funzione (13):

$$(2\,\pi\;\sigma_{\varepsilon}^{2})^{-n/2}\;\; |V_0|^{-1/2}\; \exp\{-(\,y_0 - X_0\,\beta\;)^{T}\,V_0^{-1}\;(\,y_0 - X_0\,\beta\;)/2\sigma_{\varepsilon}^{2}\,\} \tag{13}$$

(*)- Per dettagli della procedura di stima si veda Di Fonzo, 1987

3.4 – ADL(1,0)

Il modello che verrà trattato in questo paragrafo differisce da quelli sinora analizzati in quanto comprende la variabile dipendente ritardata tra le variabili esplicative: non viene considerata infatti la (7) come punto di partenza, bensì è ipotizzata una relazione più complessa, del tipo:

$$y_t = \phi y_{t-1} + \alpha + \beta x_t + \varepsilon_t \text{ , } t=1,\ldots,4n \qquad (14)$$

Nella regressione entra in gioco la variabile risposta ritardata, non presente in nessuno dei modelli esaminati fino a questo punto, mentre il disturbo è assunto *white noise*.

Dal punto di vista logico questo modello ha il pregio di non imporre una struttura stocastica agli errori, demandando al valore ritardato della variabile dipendente il compito di rappresentare l'effetto della storia passata sulla dinamica corrente del fenomeno di interesse.

Il modello è stato presentato da Salazar et al. (1997) e, con alcune differenze, da Santos Silva e Cardoso (2001).

Per dettagli di natura tecnica su questo modello e specialmente per i problemi riguardanti la fase di stima si rimanda a Di Fonzo (2003).

È importante notare che in fase di elaborazione ogni modello sarà applicato in quattro forme, ci saranno infatti le regressioni calcolate nei livelli e nei logaritmi, con e senza costante; operativamente non si tratta del miglior modo di operare, bisognerebbe infatti valutare a seconda del caso se applicare o meno la trasformata logaritmica o se inserire o meno la costante, ma per semplicità, e per avere un quadro completo, si è preferito "battere a tappeto" tutte le operazioni di stima offerte dal programma di calcolo..

4- L'ANALISI EMPIRICA

4.1- PRESENTAZIONE DEI DATI

Lo scopo della ricerca è valutare e comparare le performances di venti modelli, o meglio di 5 modelli in 4 forme diverse , ossia:

1- ADL(1,0) con costante, livelli
2- ADL(1,0) con costante, logaritmi
3- ADL(1,0) senza costante, livelli
4- ADL(1,0) senza costante, logaritmi
5- Differenza prima nei livelli con drift (Fernàndez,1981)
6- Differenza prima nei logaritmi con drift
7- Differenza prima nei livelli senza drift
8- Differenza prima nei logaritmi senza drift
9- Chow Lin con costante, livelli (AR(1))
10- Chow Lin con costante, logaritmi

11- Chow Lin senza costante, livelli

12- Chow Lin senza costante, logaritmi

13- Litterman con costante, livelli (ARIMA(1,1,0))

14- Litterman con costante, logaritmi

15- Litterman senza costante, livelli

16- Litterman senza costante, logaritmi

In appendice è possibile consultare la legenda che seguiremo da questo momento in poi per indicare i modelli utilizzati e le serie a cui ci riferiremo.

Per poter operare un confronto utilizzeremo delle procedure che ci consentiranno di disaggregare temporalmente 94 serie storiche con ognuno dei metodi sopra elencati, avendo così la possibilità di mettere a confronto i risultati ottenuti con le diverse metodologie.

Il database su cui lavoreremo è, come abbiamo già detto, lo stesso utilizzato in precedenza da Moauro e Savio; le 94 serie storiche da disaggregare sono divise fra 12 nazioni e 8 tipi di aggregati.

Le 12 nazioni sono: USA, Australia, Messico, Corea, Italia, Germania, Francia, Spagna, Olanda, UK, Giappone e Canada; in figura 1 possiamo invece vedere gli 8 aggregati e i relativi indicatori di riferimento.

	AGGREGATI	INDICATORI
1	Deflatore del Pil	Prezzi al consumo
2	Pil	Produzione industriale
3	Broad Money (M3)	Narrow Money (1)
4	Prezzi al consumo	Prezzi alla produzione
5	Importazioni CIF	Importazioni FOB
6	Tassi di interesse a breve termine	Tassi di interesse a lungo termine
7	Produzione industriale	Ordinativi industriali
8	Consumi	Pil

Figura 1- aggregati oggetto di studio e relativi indicatori di riferimento

Generalmente non è possibile valutare con esattezza la qualità di un processo di trimestralizzazione o in generale di disaggregazione temporale, dal momento che la sua necessità è conseguenza diretta di una mancanza dei dati che si cercano di stimare; nel nostro caso è differente poiché disponiamo dei dati "veri" e quindi potremo dare un giudizio fondato sull'effettivo rendimento dei modelli.

Lo scopo di questo studio è stabilire quale metodo di disaggregazione debba essere usato o almeno quale dia risultati più fedeli agli originali, con una particolare attenzione ai primi quattro modelli e alle loro performance, in quanto più recenti e qui in fase di test. Una particolare attenzione deve essere posta al tredicesimo modello della lista: Chow-Lin con costante, nei livelli. Questo metodo è infatti utilizzato abitualmente non solo dall'ISTAT, ma anche da numerosi enti statistici, in particolare nel resto d'Europa ricordiamo Francia, Spagna e Portogallo..

Il lavoro vede il suo primo passo in una fase di stima: utilizzeremo a questo scopo il software Gauss 4.0, appoggiandoci al programma DynChow[1] ancora in fase sperimentale.

Costruiremo quindi 5 piccoli programmi che avranno il compito di disaggregare coi 5 metodi principali in ognuna delle 4 forme tutte le 94 serie storiche, producendo così 470 output contenenti i dati stimati.

Il secondo passo sarà riunire tutti gli output in 94 tabelle Excel assieme ai dati reali, potendo così operare un confronto basato sugli errori, errori assoluti , varianza e diversi altri indicatori. A questo punto sarà possibile costruire tabelle e grafici riassuntivi che spieghino in maniera chiara i risultati, e per questo scopo utilizzeremo principalmente Excel ed EViews 4.0.

4.2 – PROBLEMI RISCONTRATI

Com'era prevedibile non tutte le elaborazioni hanno dato i risultati sperati, infatti già dalle prime fasi si sono incontrati alcuni problemi; nella fase di confronto sono emerse delle difficoltà, alcune irrilevanti ai fini dei

[1] - DynChow è un programma di calcolo sviluppato da Eurostat, sotto la supervisione di Gian Luigi Mazzi e Giovanni Savio.

risultati, altre invece hanno portato ad una riduzione dei dati piuttosto consistente. Tutti i dati disaggregati riguardanti i consumi trimestrali, ossia i file qcg, hanno mostrato seri problemi di affidabilità degli indicatori, per cui si è preferito escluderle dall' analisi. Ci troviamo quindi a svolgere l'analisi su un database di 82 serie disaggregate con 16 procedure.

Lo studio non perde comunque il suo interesse dal momento che il database rimane molto consistente.

4.3 – GLI INDICATORI DI PERFORMANCE

Il principale indicatore di qualità delle stime che utilizzeremo è il *Root Mean Squared Percentage Error* (RMSPE):

$$\left(\sqrt{\sum_{i=1}^{n}\left(\frac{y_i-\hat{y}_i}{y_i}\right)^2}\ /n\right)\text{x } 100 \qquad (15)$$

Questo è un indice molto importante dal momento che ci permette di calcolare per ognuno dei modelli in tutte le serie disaggregate una misura di bontà delle stime confrontabile fra le serie stesse; in effetti non è altro che la media quadratica degli errori relativi .

Oltre a costituire un indicatore semplice ed efficace l'RMSPE è ciò su cui si basano Moauro e Savio (2004), che prendono come punto di riferimento il modello da loro presentato, il SUTSE, per calcolare il rapporto fra l'RMSPE di questo e tutti gli altri; dalle loro elaborazioni il SUTSE risulta essere senza ombra di dubbio il migliore dei modelli. I risultati ottenuti nel lavoro qui presentato differiscono in maniera piuttosto significativa da quelli di Moauro e Savio: circa il 25% dei risultati è diverso nei due studi e comunque i dati che produrremo passeranno per un processo di scrematura che nel lavoro di Moauro e Savio non esisteva.

5- RISULTATI

5.1- PRIMA ANALISI

Lo studio svolto produce una gran quantità di dati, che non danno un'informazione chiara e univoca che ci possa condurre ad una scelta immediata di quale sia il migliore dei modelli presi in considerazione.

Molto dipende da quanti e quali dati vengono scelti per rappresentare i nostri risultati: inizialmente ci occuperemo di un'analisi basata sullo studio della totalità dei dati, senza operare alcun tipo di scrematura.

Il miglior indicatore a nostra disposizione è l'RMSPE, ed è quello sul quale baseremo la maggior parte della nostra analisi. Per ognuna delle 82 serie prese in considerazione abbiamo calcolato l'RMSPE di tutti i modelli, di modo da poter valutare il modello che da i risultati più attendibili in ciascuna serie, ossia quello col valore minore, per poter stilare una classifica .

In figura 1 possiamo vedere quante volte in percentuale ognuno dei modelli ha prodotto le stime migliori.

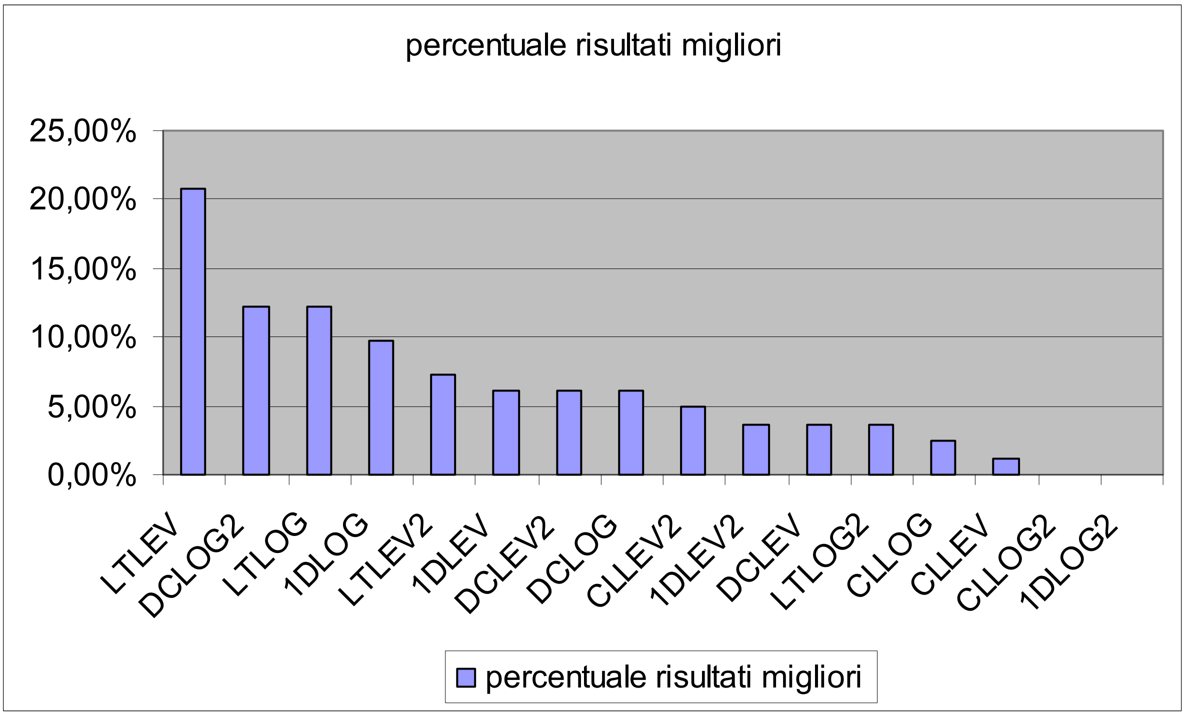


figura 1

Il modello ltlev sembra essere il più efficace, dal momento che nel 20% dei casi ha prodotto le stime più vicine alla realtà, ma vedremo in seguito come questo risultato sia ancora parziale.

Prendiamo in considerazione i risultati del punto precedente raggruppati per i quattro modelli fondamentali

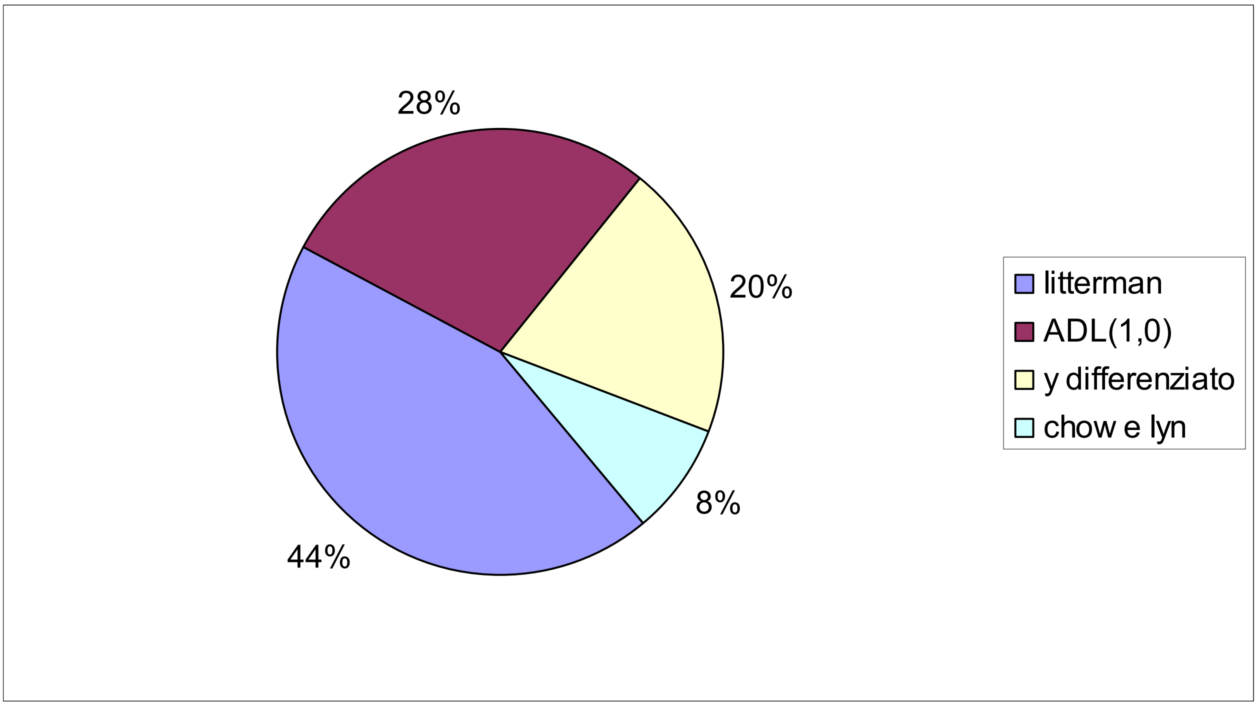
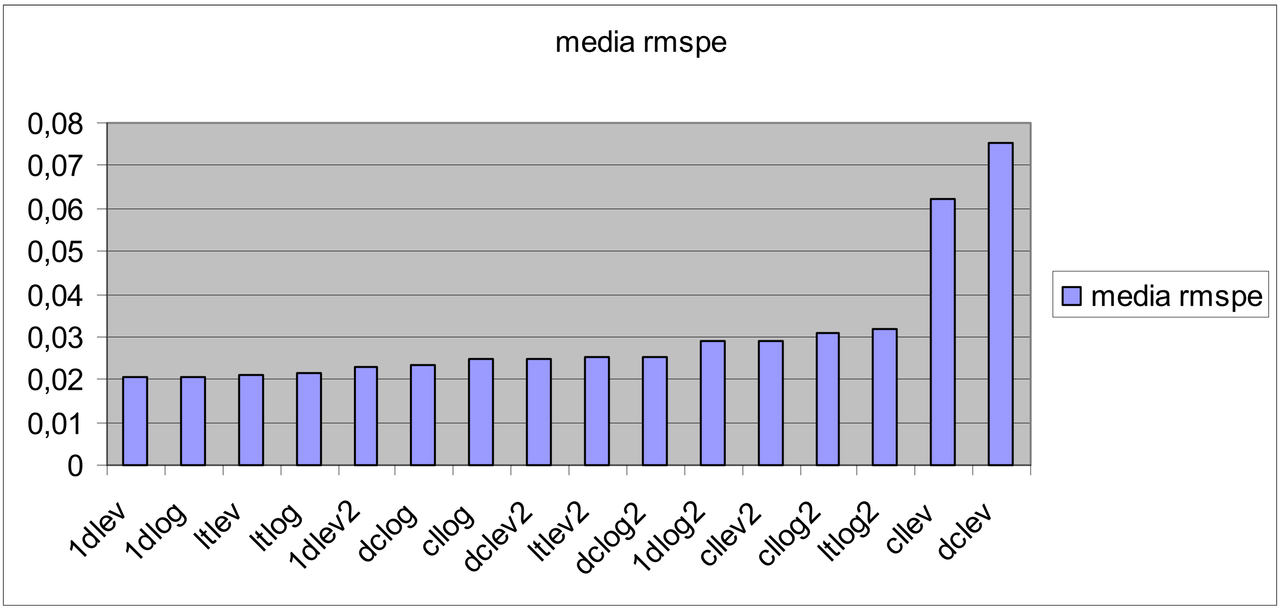


Figura2

Da questa analisi iniziale spiccano due cose:una prevalenza dei modelli di Litterman sugli altri e una inefficienza piuttosto marcata dei modelli Chow-Lin che in tutte le loro forme non si comportano meglio della “concorrenza” in nemmeno un decimo dei casi; questo fatto risulta molto significativo considerando che cllev, ossia Chow Lin nei livelli con costante è il metodo ufficiale utilizzato dall’ISTAT .

A questo punto andiamo a calcolare media e mediana degli RMSPE di tutte le serie per ognuno dei modelli e stiliamo una classifica che le ordini dalla più piccola alla più grande, questo dovrebbe ordinare i modelli dal migliore al peggiore data la natura dell’indicatore. In figura 3 possiamo vedere i due ordinamenti: quello per media in alto e quello per mediana in basso.



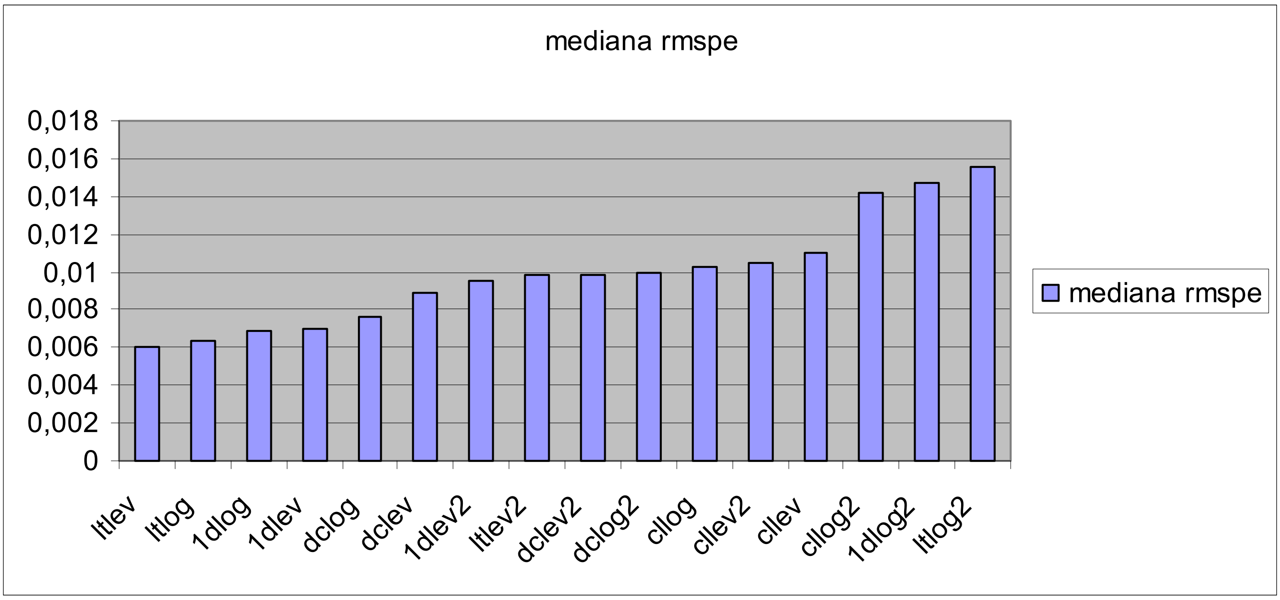


Figura3

Si può notare come la classifica per media coincida solo a grandi linee con quella per mediana, a riprova di una diversa distribuzione dei valori intorno alla media a seconda del modello preso in considerazione.(Dei valori di dclev e cllev nel primo grafico, visibilmente più alti degli altri, parleremo dopo.)

Un esempio ci è fornito dal modello 1dlev , ossia Fernàndez, se raffrontato con ltlev, ossia Litterman a livelli con costante.

Qui sotto prendiamo in considerazione delle tabelle descrittive che riguardano questi due modelli, rispettivamente 1dlev e ltlev.

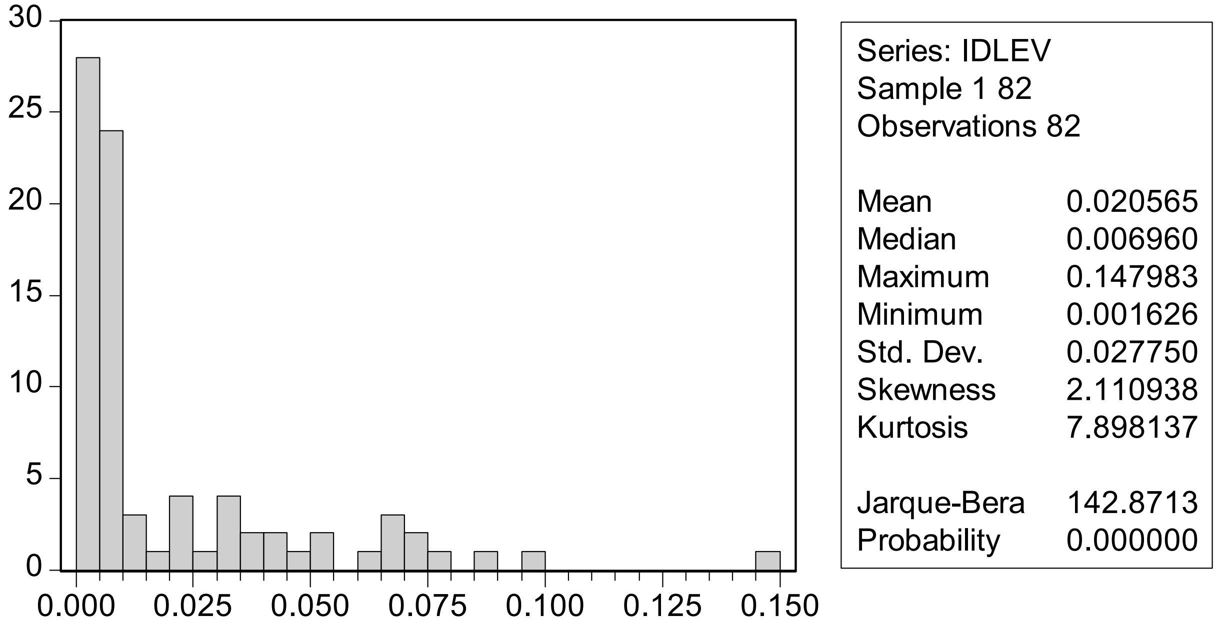


figura 4- tabella descrittiva dell'andamento dell'RMSPE del modello 1dlev

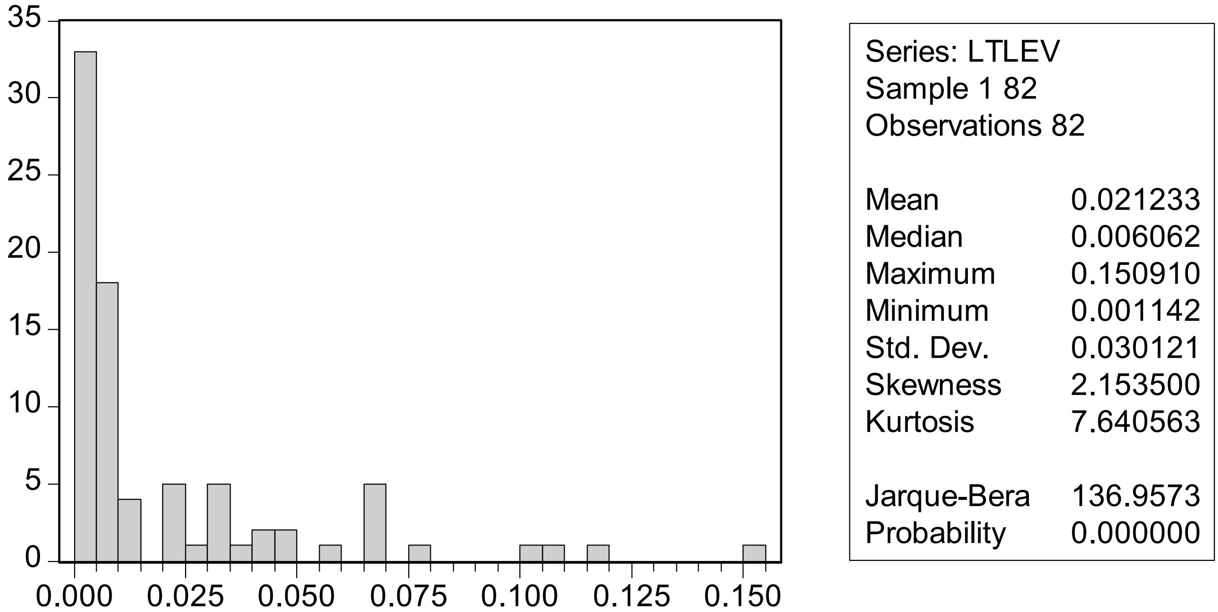


figura 5- tabella descrittiva dell'andamento dell'RMSPE del modello ltlev

Questi due modelli sono i migliori per media (1dlev) e per mediana (ltlev).

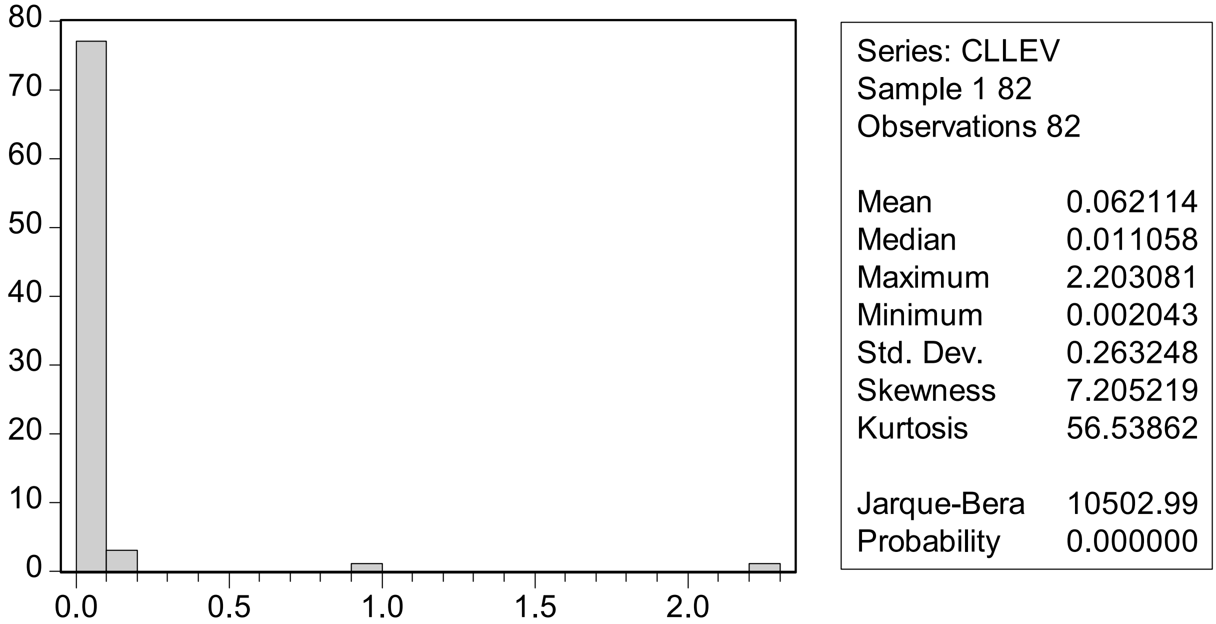
1dlev ha media più bassa perché, come si può vedere dal grafico, ha meno

valori estremi che tendono a spostare la media verso l’alto.

Da questi dati potremmo iniziare a dare dei giudizi complessivi sui modelli ma come vedremo dal paragrafo successivo sarebbe decisamente prematuro.

5.2-VALORI ANOMALI E FASE DI SCREMATURA

La situazione in cui i valori estremi influenzano fortemente il valore medio è riscontrabile nei due modelli che nel primo grafico in figura 3 mostrano un RMSPE molto alto, ossia dclev e cllev. Vediamo perché :



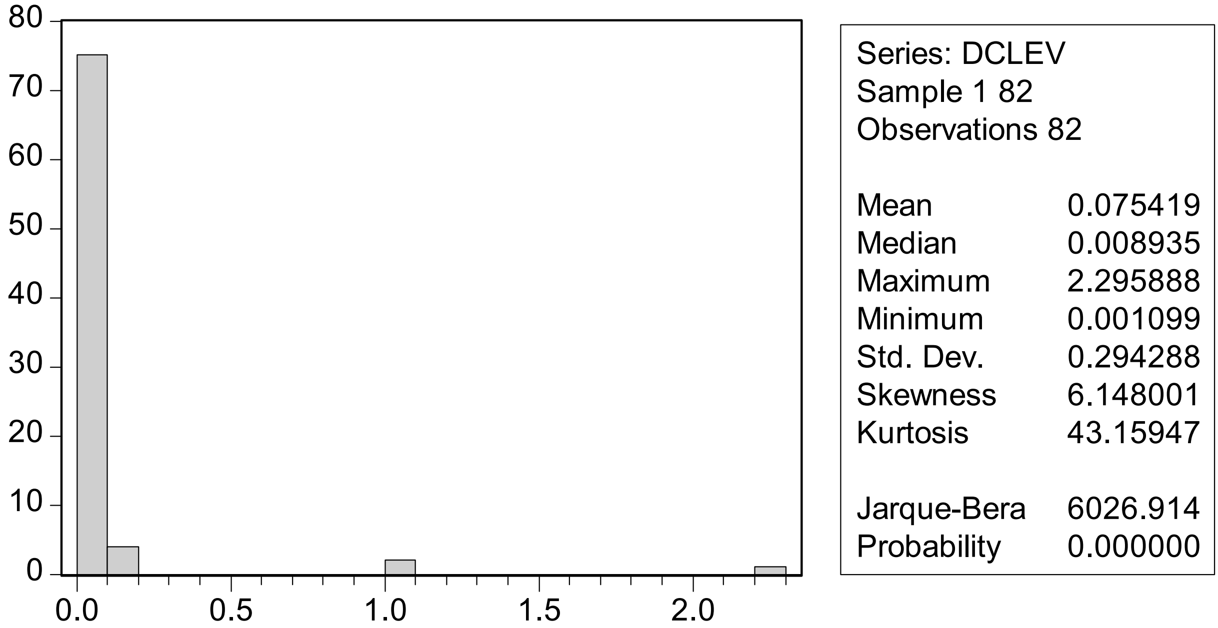


figura 6- tabelle descrittive dell'andamento dell'RMSPE dei modelli cllev e dclev

Questi modelli sono i peggiori per quel che riguarda l'RMSPE medio, ma è evidente dalla figura 6 che la causa di questo non è attribuibile ad un generale malfunzionamento dei modelli bensì alla presenza di pochi valori molto alti dell'RMSPE che alzano in maniera drastica la media; fatto confermato dalla migliore posizione in classifica per quel che riguarda la mediana.

Questa situazione denota in maniera piuttosto chiara che un'analisi "cieca" dei dati presi in considerazione a questo punto fornisce una visione parziale della bontà dei modelli; risulta quindi evidente che prima di poter confrontare i dati bisogna depurarli dalla presenza di valori anomali o quantomeno cercare di capire se la causa di questi è insita nella natura dei modelli o se è dovuta ad errori di altro genere.

Vediamo in figura 6 che il valore massimo raggiunto dall'RMSPE del modello dclev è 2,295 mentre per cllev è 2,203 ; esaminiamo quindi la serie in cui ci troviamo ad avere questo risultato, che è la stessa per tutti e due i modelli, ossia la serie kormnb.

Qui sotto (figura 7) possiamo vedere l'andamento della serie kormnb nei primi 5 anni disaggregata mensilmente col metodo dclev, col metodo cllev (figura 8) e col metodo ltlev (figura 9).

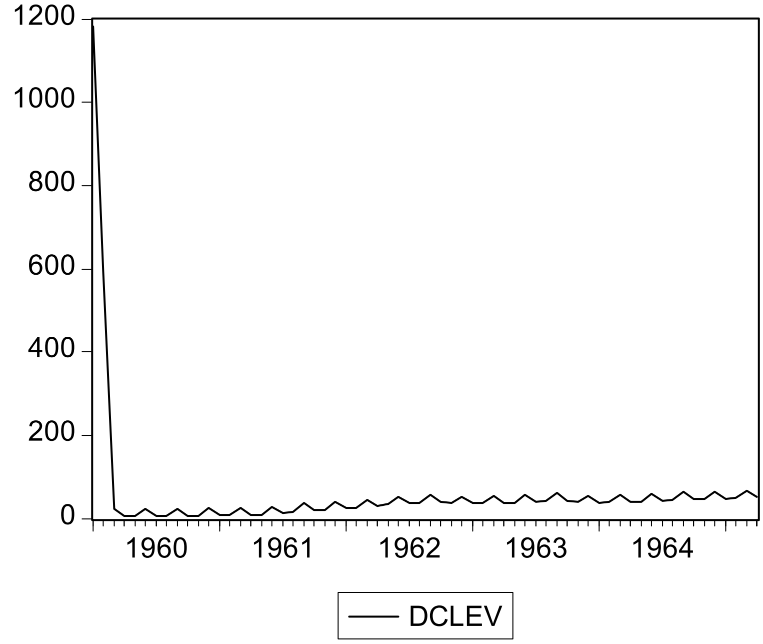


figura 7-andamento della serie kormnb disaggregata col metodo dclev

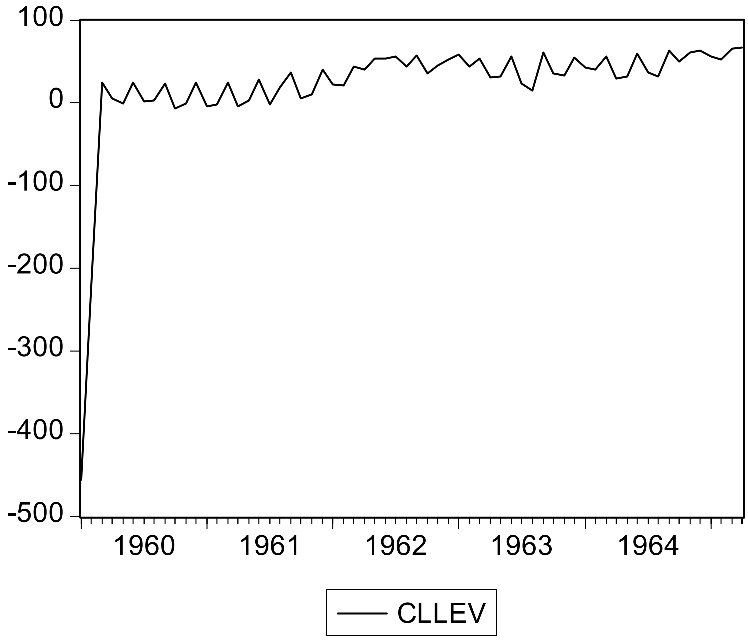


figura 8-andamento della serie kormnb disaggregata col metodo cllev

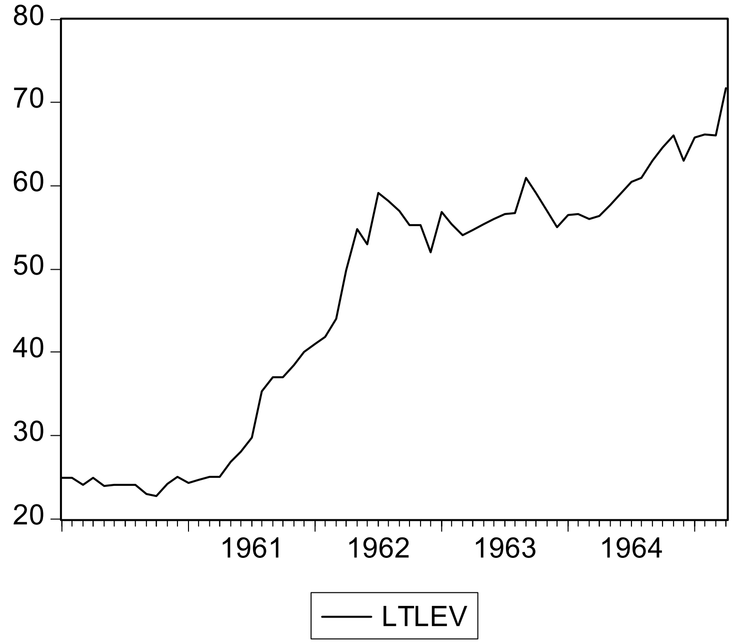


figura 9-andamento della serie kormnb disaggregata col metodo ltlev

Basta un’ occhiata per capire che qualcosa non quadra: in tutte le serie c’è un andamento crescente con la serie che da valori prossimi allo zero all’inizio del 1960 si sposta verso valori prossimi al cento alla fine del 1964, ma i primi valori dati dai modelli dclev e cllev sono visibilmente irrealistici, addirittura il modello cllev nei dati del 1960 stima valori negativi, che non hanno alcun senso economico.

Questo è un problema raro ma tutt’altro che irrilevante: il programma usato per fare le stime talvolta da degli output che nei primi valori sono sostanzialmente anomali, al punto che quei pochi dati “traviano” l’intera analisi dei modelli. Questa situazione può essere causata da errori di programmazione o da errori insiti nei modelli; gli output in questione sono comunque inservibili e debbono essere eliminati.

Prima di poter avere dei risultati veramente significativi bisogna allora procedere con un processo di scrematura dei dati per il quale terremo conto di queste regole:

1- Eliminare i dati palesemente anomali
2- eliminare le serie che sembrano non coincidere con i dati originali (vincolo di aggregazione non rispettato o RMSPE troppo alti per tutti i modelli)
3- eliminare le serie disaggregate in cui il parametro AR stimato sia minore di zero per evitare il rischio di avere una distribuzione fortemente variabile delle discrepanze
4- eliminare tutte le serie disaggregate il cui modello abbia prodotto un phi pari a 0,99

Qui entra in gioco una componente empirica, dal momento che bisogna riconoscere quando i valori in output sono illogici; si tratta di un passaggio delicato che pone qualche problema al quale si è cercato di ovviare secondo criteri di buon senso.

Gli ultimi due punti indicano regole consigliabili ma non strettamente necessarie, ed è un argomento piuttosto delicato nel nostro studio poiché è una situazione che si ripete spesso. Procederemo quindi con due distinte analisi: nella prima prenderemo in considerazione i primi due punti sopra descritti e opereremo una prima scrematura dei dati, nella seconda terremo conto anche del terzo e del quarto punto.

5.3-I FASE DI SCREMATURA : STUDIO SU 75 SERIE DISAGGREGATE

Il nostro database consta di 82 serie disaggregate, sette di queste vengono eliminate in questa prima fase di scrematura e vengono eliminate per ognuno dei modelli, dal momento che togliere solamente i valori dei modelli che riscontrano i problemi indicati ai primi tre punti della pagina precedente li favorisce troppo nei confronti della concorrenza .

Le sette serie che vengono eliminate sono :

- mexqdp – valori negativi in dclev e RMSPE alti per tutti
- kor e mex mnb – valori negativi o prime stime anomale in diversi modelli dc e cl
- itamls – comportamento anomalo dei modelli dc
- mexmls – vincolo di aggregazione non rispettato
- itamid – comportamento anomalo dei modelli dc

Per ognuna delle 75 serie oggetto di studio calcoliamo l'RMSPE di ognuno dei modelli e vediamo quante volte ogni modello ha rendimento migliore, ossia RMSPE minimo, in percentuale sul totale.

I risultati sono riportati in figura 10

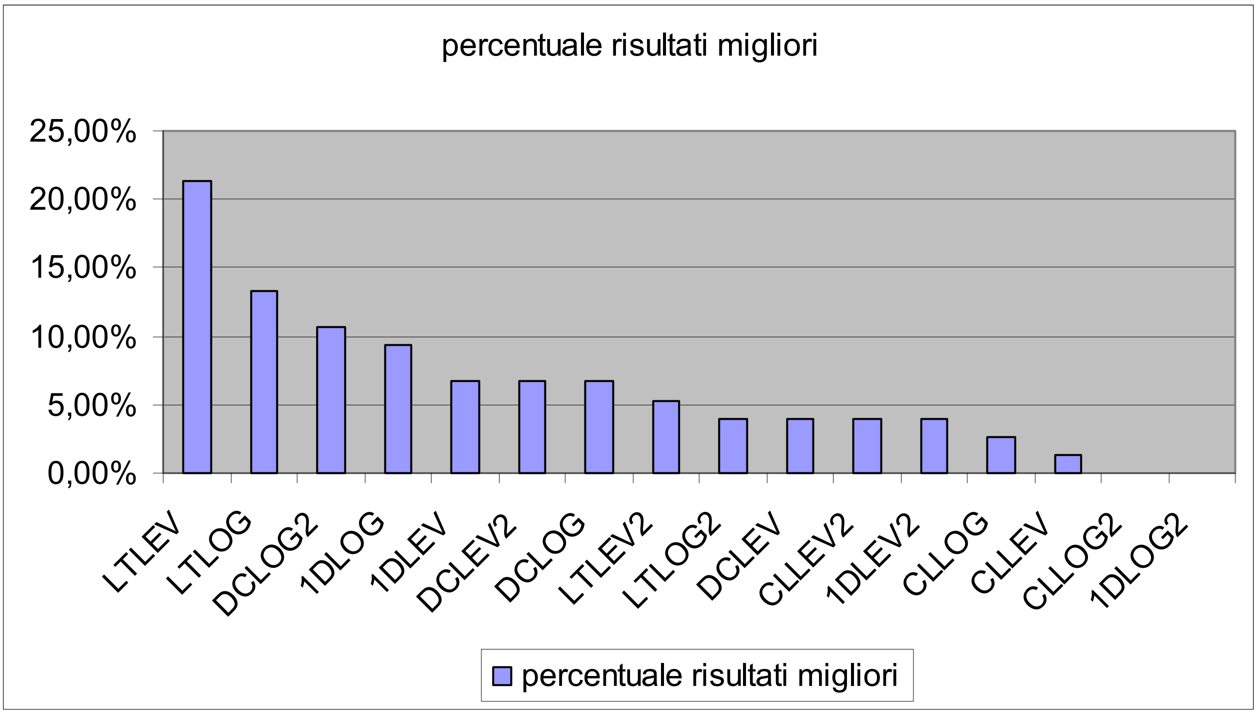


figura 10

Se dividiamo le nostre osservazioni raggruppando i modelli logaritmici e non e quelli con e senza costante nei 4 principali abbiamo i risultati in figura 11.

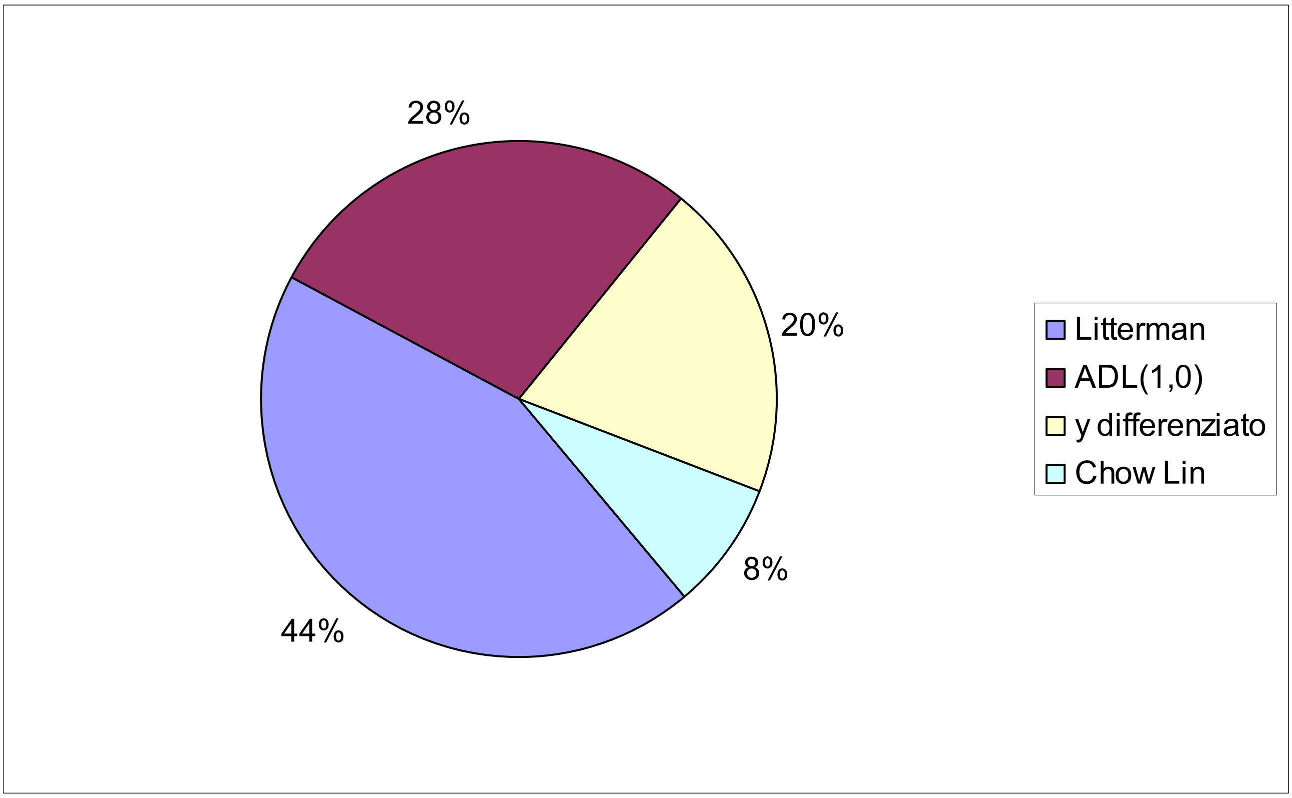
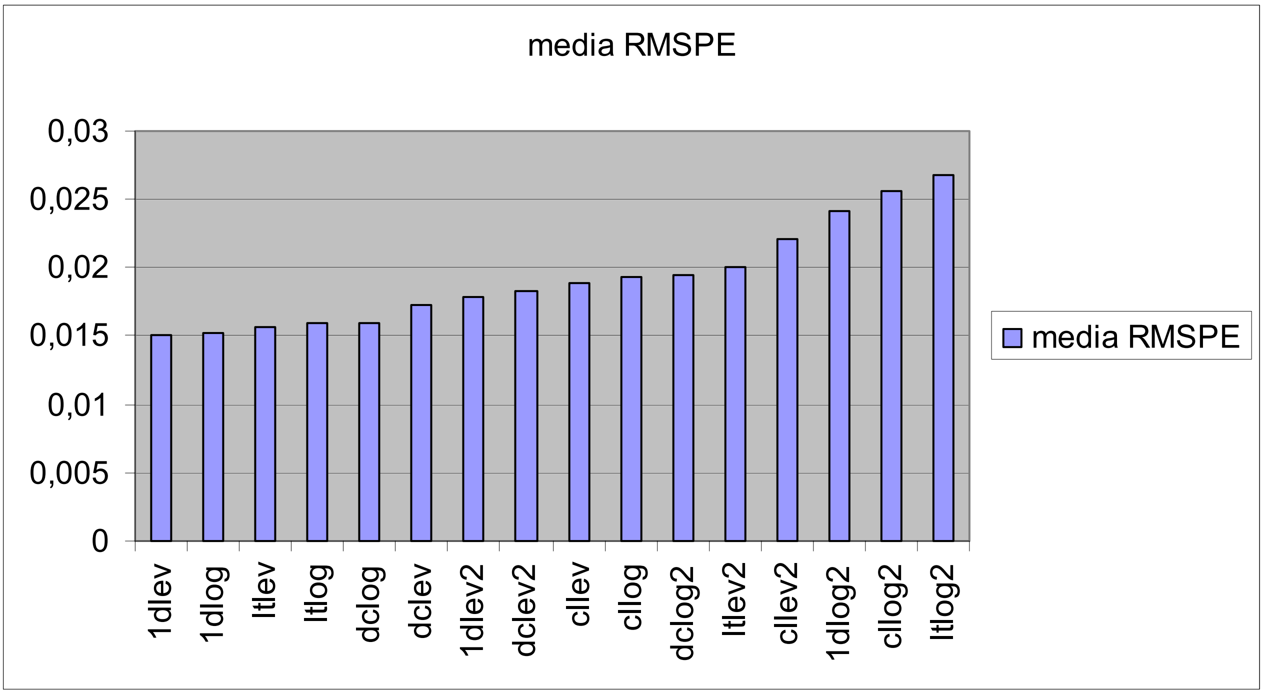


figura 11

I risultati in figura 10 e 11 sono piuttosto simili a quelli ottenuti prima della scrematura dei dati, grazie al basso numero di serie e all'eterogeneità dei risultati eliminati.

Da questi due grafici appaiono evidenti due informazioni: la prima è una prevalenza dei modelli di Litterman, in particolare dei modelli con costante logaritmici e non, su tutti gli altri, la seconda è lo scarso rendimento dei modelli Chow Lin, i quali nemmeno nel 10% dei casi si impongono sugli altri.

Andiamo ora ad osservare media e mediana degli rmspe di tutti i modelli.



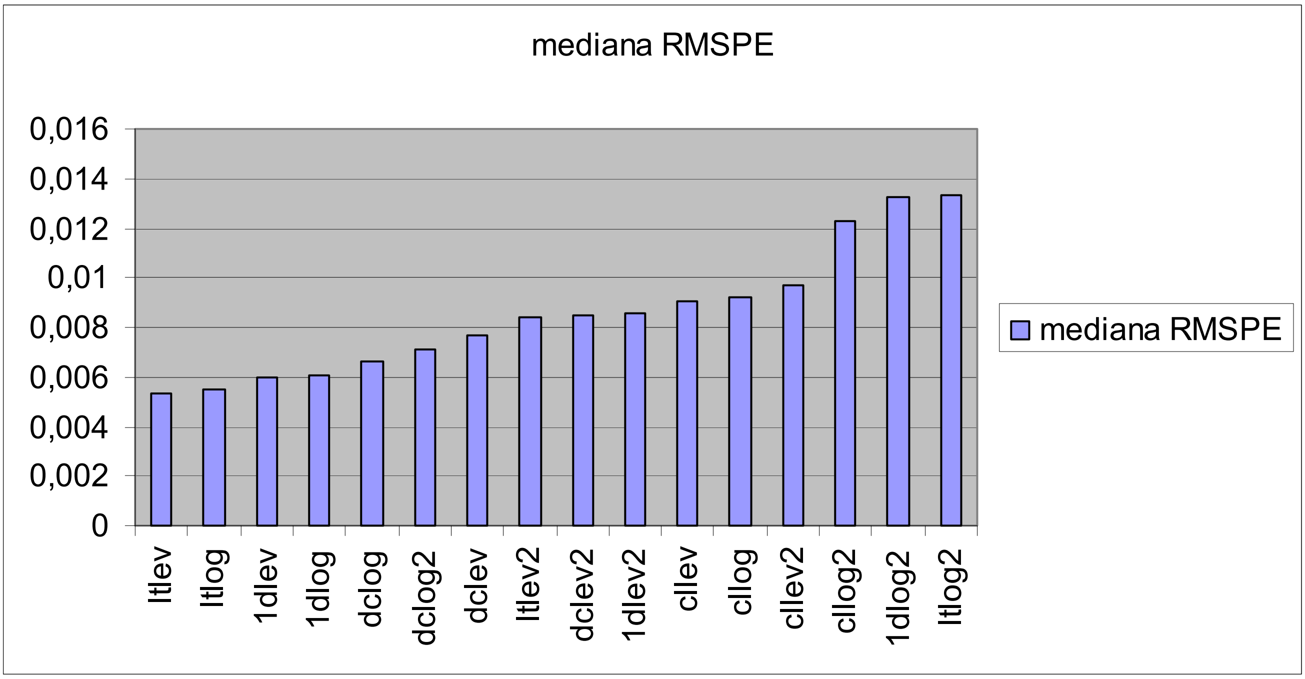


figura 12 e 13 –media e mediana degli rmspe dopo la pulizia dei dati

Dalle figure 12 e 13 emerge una situazione diversa rispetto a quella precedente alla scrematura,anche se non nelle prime 4 posizioni; i migliori 4 modelli sono quelli di Litterman con costante, nei livelli e nei logaritmi, e quelli con y differenziato con costante, nei livelli (Fernàndez,1981) e nei logaritmi. Non ci sono sostanziali differenze fra il comportamento dei vari modelli se si guarda la media o la mediana, e ciò significa che tutti i modelli hanno una distribuzione dei valori dell'RMSPE simile intorno alla media, e questa distribuzione è in tutti i casi sbilanciata a sinistra della media dal momento che i valori della mediana sono sempre molto più piccoli di quella della media.

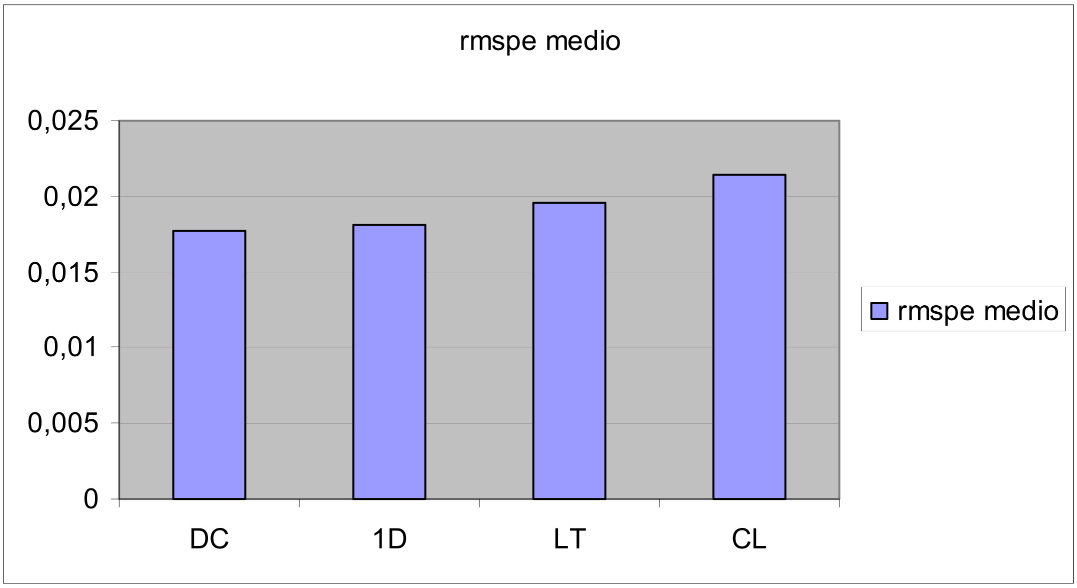


figura 14-rmspe medio per modelli raggruppati

Dalla figura 14 possiamo vedere l'RMSPE medio per le quattro principali categorie di modelli. I valori sono più livellati rispetto ai grafici già visti e colpisce che i modelli dc nell'insieme siano i migliori dal momento che non sembravano eccellere in figura 12 e 13; questo avviene perché nessun modello dc è realmente "debole" come lo sono tutti gli altri se presi senza costante e calcolati sui logaritmi ed è indice di una maggiore adattabilità dei modelli dc ai dati.

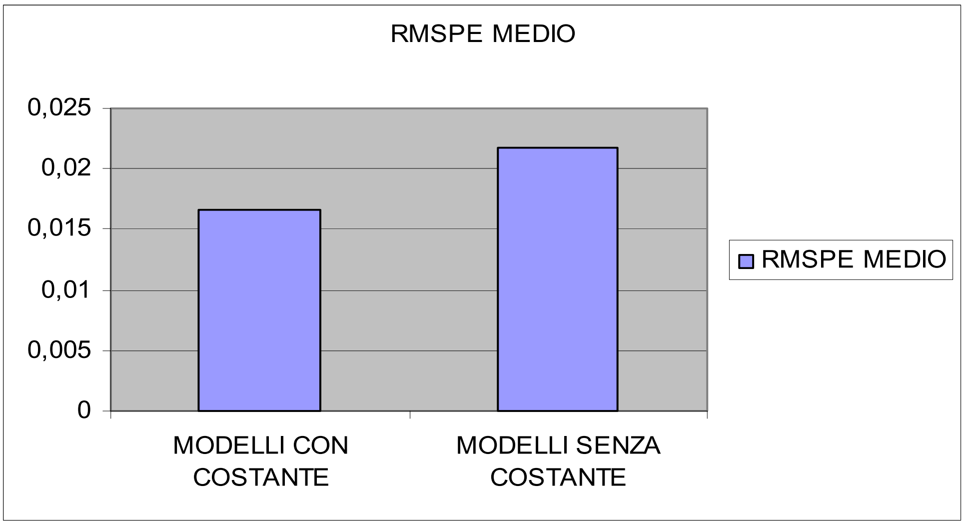


figura 15-rmspe medio per modelli con costante(sinistra) ed rmspe medio per modelli senza costante

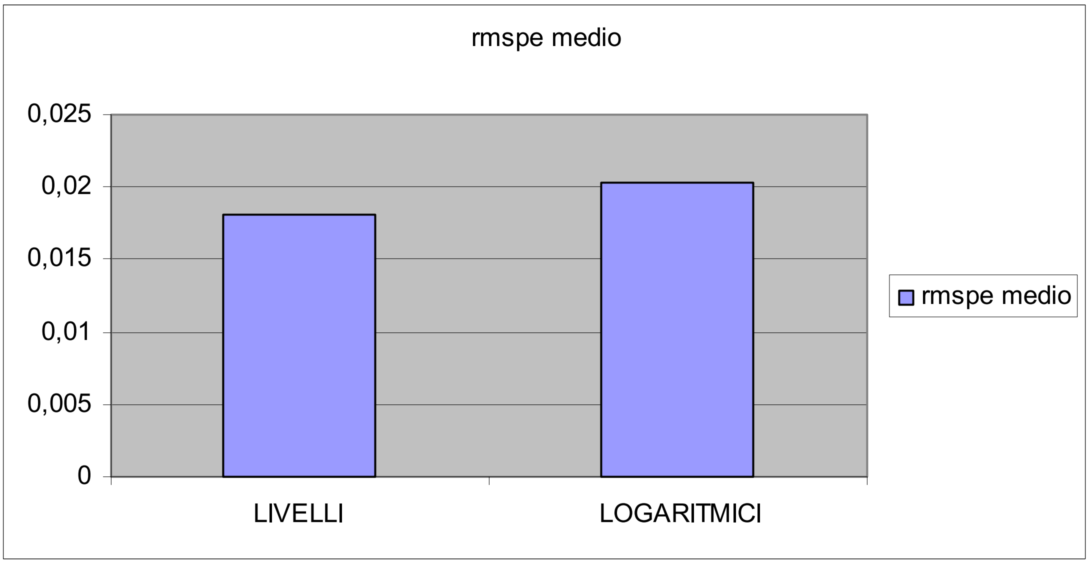


figura 16-rmspe medio per modelli calcolati sui livelli ed rmspe medio per modelli logaritmici

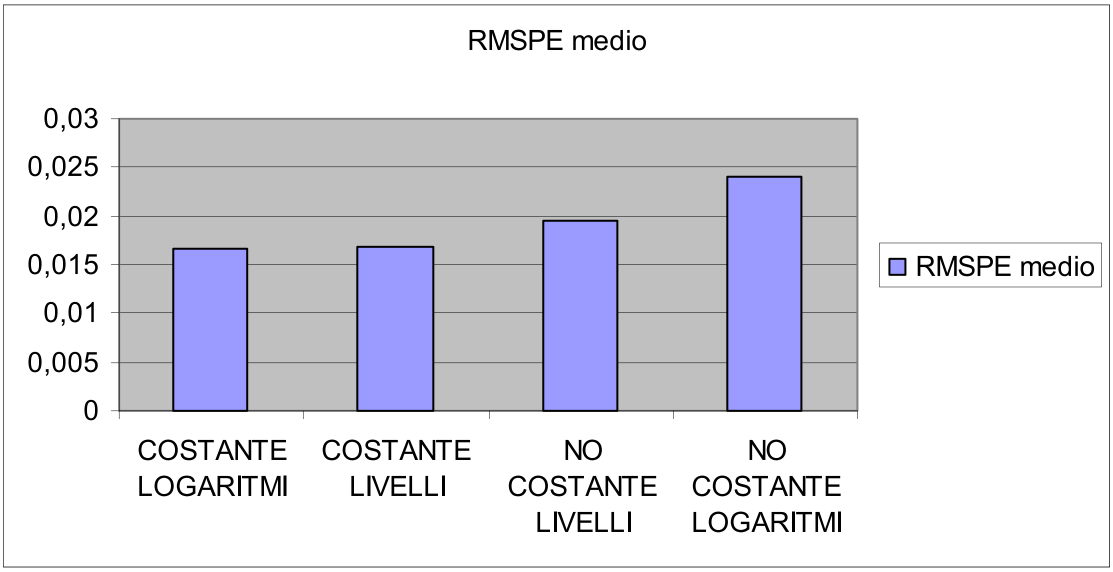


figura 17-rmspe medio calcolato per 4 diverse tipologie di modelli

In figura 15 e 16 possiamo vedere che i modelli con costante danno mediamente risultati migliori di quelli senza costante e che i modelli calcolati sui livelli danno migliori risultati di quelli calcolati sui logaritmi, diventa quindi curioso il dato rappresentato in figura 17 che ci mostra come siano i modelli con costante calcolati sui logaritmi ad avere RMSPE medio minore, anche se con un risultato quasi identico a quello dei modelli con costante calcolati sui livelli. Questo dato è inaspettato, ma significa solamente che i modelli logaritmici hanno RMSPE medio molto alto per il contributo pesantemente negativo dei modelli senza costante.

Proviamo ora a vedere se c'è connessione tra il tipo di aggregato stimato e la precisione dei modelli nella stima: a questo scopo prendiamo i 4 modelli calcolati sui livelli con costante e vediamo qual è stata la loro resa media in termini di RMSPE per ognuno degli aggregati che abbiamo studiato.

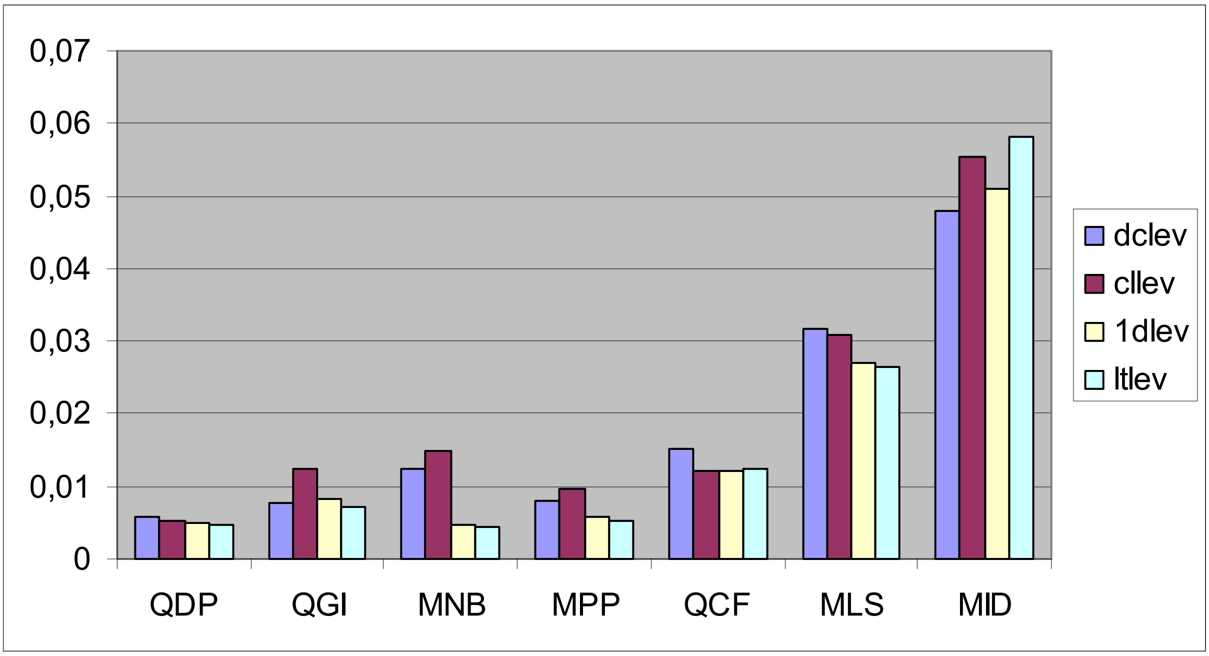


figura 18- rmspe medio dei 4 modelli indicati a destra per ognuno dei sette aggregati oggetto di studio

Dalla figura 18 risulta un legame piuttosto forte fra l'aggregato da stimare e la precisione con cui la stima avviene: possiamo vedere come gli RMSPE medi dell'aggregato qdp abbiano un valori pari indicativamente ad un decimo di quelli dell'aggregato mid; per i dodici modelli non inclusi in figura 18 la situazione è analoga.

Se guardiamo il rendimento dei modelli suddiviso per nazione ci troviamo di fronte ad una situazione simile: la media degli RMSPE varia di nazione in nazione in maniera univoca fra i vari modelli (figura 19)

Lascia piuttosto perplessi la situazione sviluppatasi per la Germania, possiamo infatti vedere come per questa nazione le stime migliori siano state date dai modelli dc e cl, addirittura le migliori in assoluto da cllev.

Come in figura 18 sono stati omessi gli altri dodici modelli per i quali la situazione è comunque analoga.

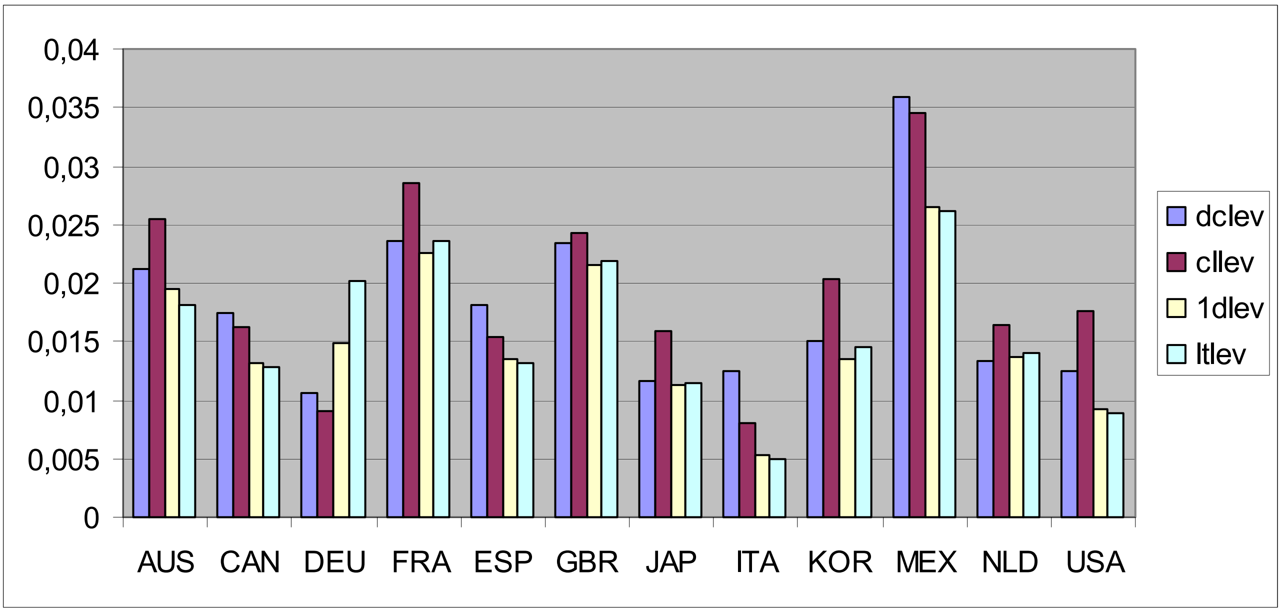
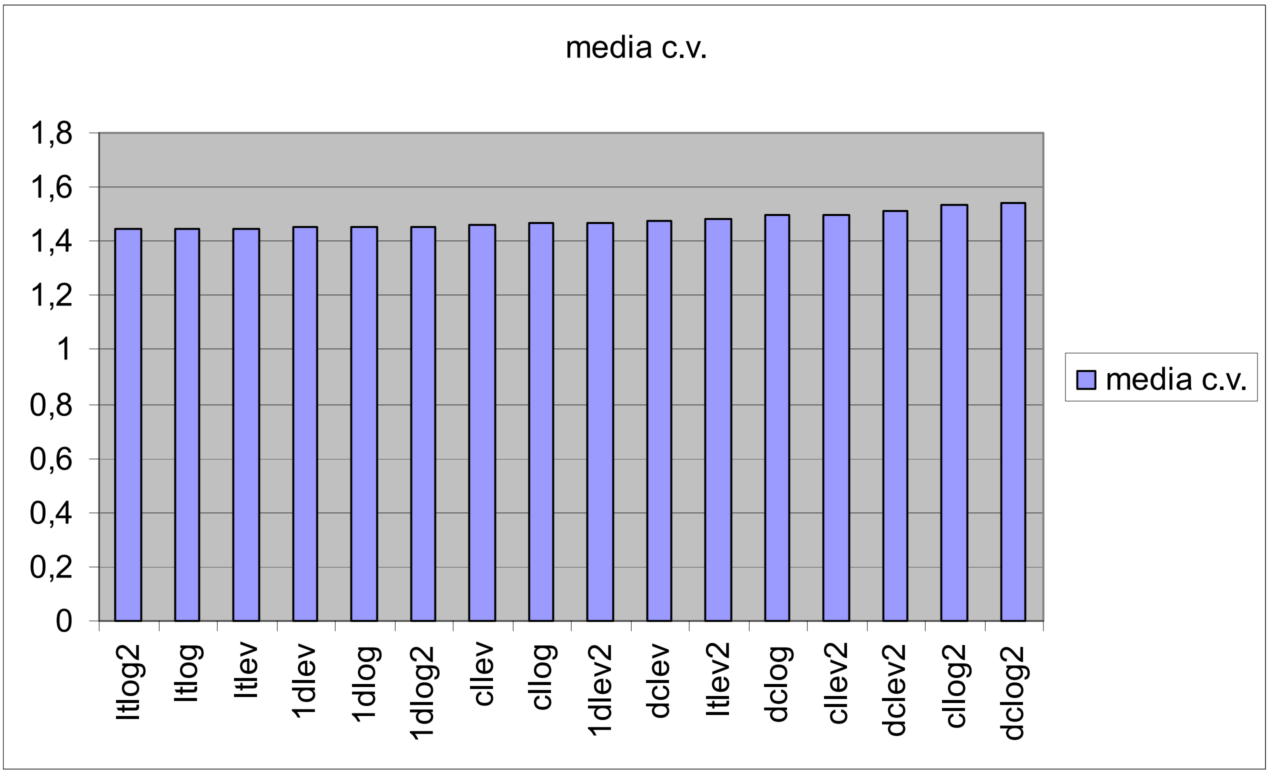
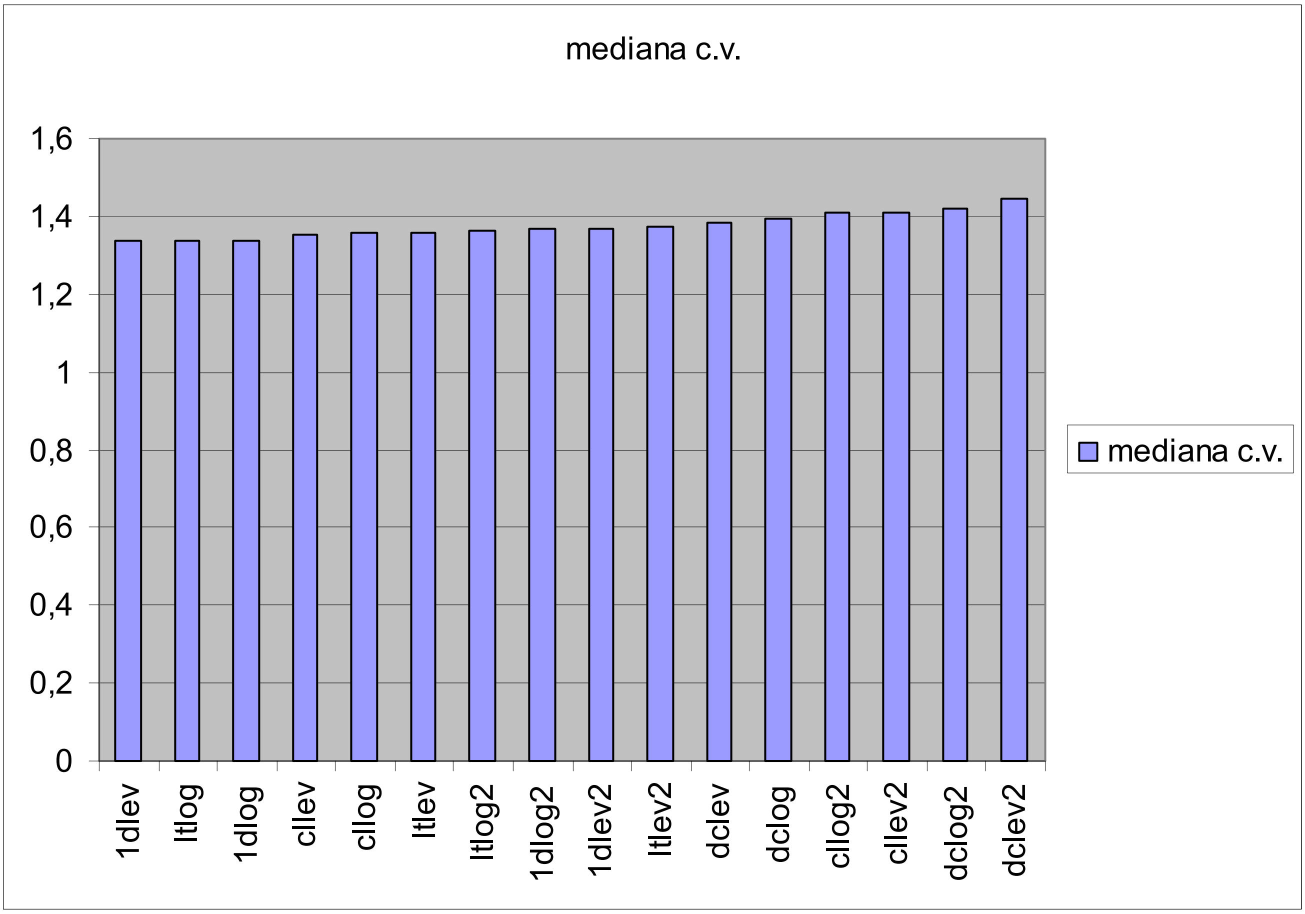


figura 19- rmspe medio dei 4 modelli indicati a destra per ognuna delle 12 nazioni oggetto di studio

Alla luce di questi due grafici non desta stupore sapere che il valore più basso di RMSPE in tutto il database finora studiato si trovi nella serie disaggregata itampp, dal momento che l'aggregato mpp ha valori bassi di RMSPE così come la nazione Italia.

Controlliamo ora il coefficiente di variazione per avere una misura confrontabile della varianza degli errori dei vari modelli.


media c.v.
1,8
1,6
1,4
1,2
1
0,8
0,6
0,4
0,2
0
ltlog2
ltlog
ltlev
1dlev
1dlog
1dlog2
cllev
cllog
1dlev2
dclev
ltlev2
dclog
cllev2
dclev2
cllog2
dclog2
media c.v.


mediana c.v.
1,6
1,4
1,2
1
0,8
0,6
0,4
0,2
0
1dlev
ltlog
1dlog
cllev
cllog
ltlev
ltlog2
1dlog2
1dlev2
ltlev2
dclev
dclog
cllog2
cllev2
dclog2
dclev2
mediana c.v.

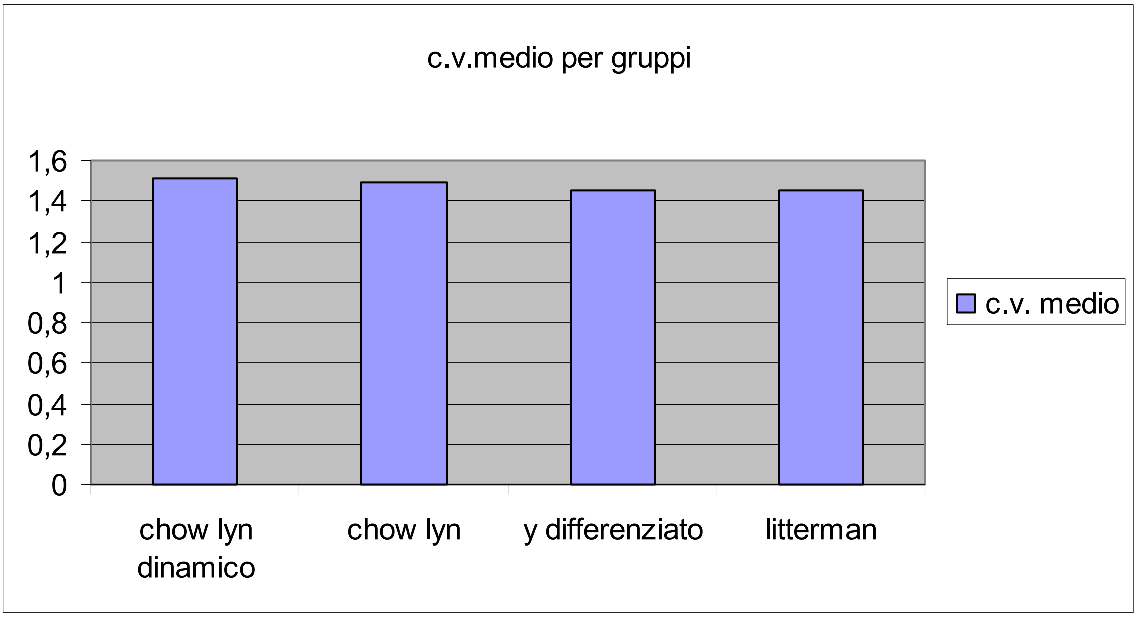


figura 20- media e mediana dei coefficienti di variazione di tutti i 16 modelli e dei 4 modelli principali

Come si può vedere le differenze in termini di variabilità tra i modelli sono minime. La classifica dei coefficienti di variazione segue più o meno le precedenti basate sugli RMSPE se non fosse per il dato che vede in testa Litterman senza costante calcolato sui logaritmi, che sinora si è sempre dimostrato un modello dai rendimenti scadenti. Questo fatto è indice di una variabilità standardizzata bassa per questo modello, ma si tratta comunque di una variabilità intorno ad un valore medio molto alto

5.4-II FASE DI SCREMATURA

5.4.1-E' REALMENTE NECESSARIO PROCEDERE AD UN' ULTERIORE SELEZIONE DEI DATI?

In questo capitolo analizzeremo le serie rimanenti dopo una scrematura completa seguendo tutti i 4 punti elencati a pagina 29.

Questo vuol dire che dalle 75 serie analizzate al paragrafo precedente toglieremo tutte le serie disaggregate da modelli caratterizzati da phi=0,99 o minore di zero. Non toglieremo però le serie disaggregate con tutti i metodi ma solamente con quelli afflitti dal problema in questione.

Ci troveremo quindi a lavorare su un database per cui ognuno dei 16 modelli sarà utilizzato al massimo in 75 serie, nel caso in cui non sia mai afflitto dal problema sopra descritto.

Sono soprattutto i modelli di Litterman a produrre phi pari a 0,99 e in numero molto elevato di serie, i casi di phi negativo sono invece solamente tre su un totale di 1200 (16*75) serie disaggregate, e sono sempre per i modelli di Lietterman(senza costante).

Qui sotto possiamo vedere che percentuale di perdita dei dati ha riportato ognuno dei modelli.

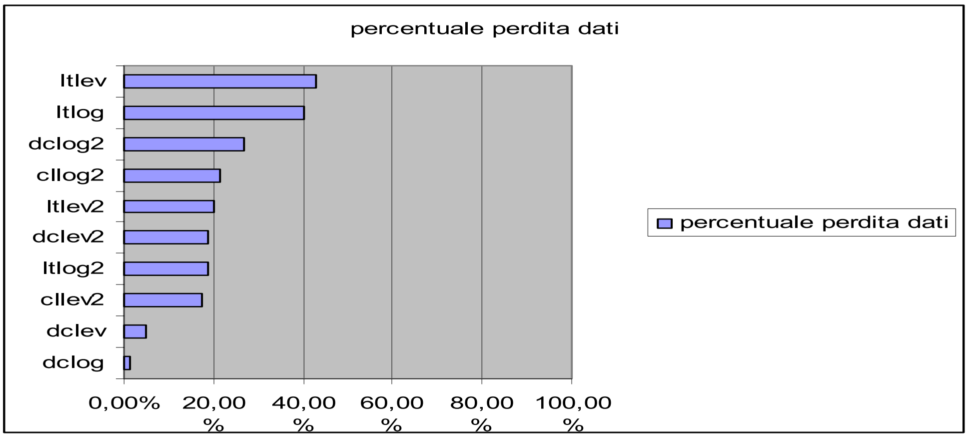


Figura 21- percentuale di serie disaggregate con phi pari a 0,99. I modelli non indicati non presentano questo problema in nessuna delle serie studiate.

Dalla figura 21 vediamo come i modelli di Litterman con costante danno luogo ad una perdita di dati drammatica, intorno al 40% .

Anche tutti i modelli senza costante, a parte quelli nella variabile y differenziata che non hanno questo problema, sono afflitti da questo problema in maniera significativa, con caduta intorno al 20%.

Operativamente un modello che produca stime non ammissibili nel 20 o peggio nel 40 % dei casi trattati non è certamente buona cosa, si tratta di percentuali troppo alte che rendono questi metodi a livello pratico inutilizzabili.

Proviamo quindi a capire se è realmente necessario operare questo taglio dei dati vedendo se e quanto è negativo l’effetto dei modelli con phi a 0,99 o negativo sul totale.

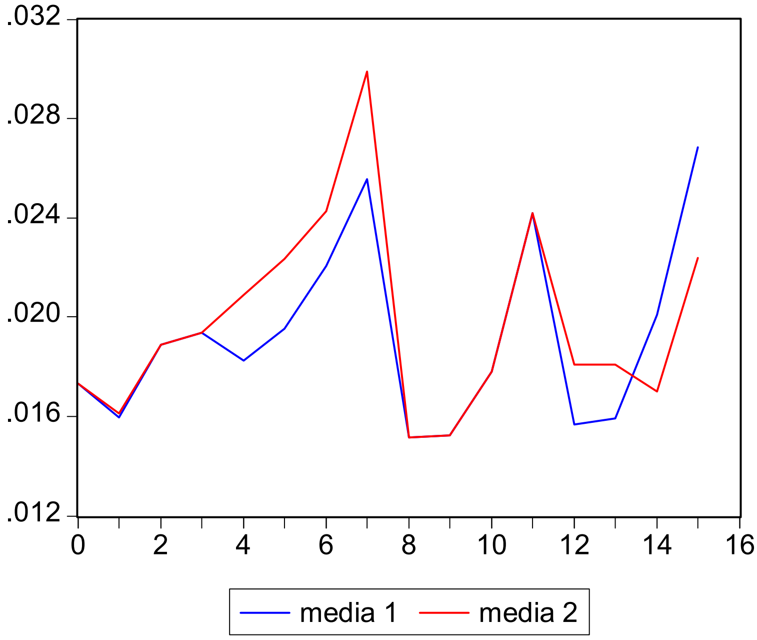


figura 22- grafico che rappresenta in ordinata il valore medio dell'RMSPE per ognuno dei 16 modelli in ascissa, nel caso in cui si tengano i modelli con phi a 0,99 (in blu) o nel caso in cui si eliminano (in rosso).

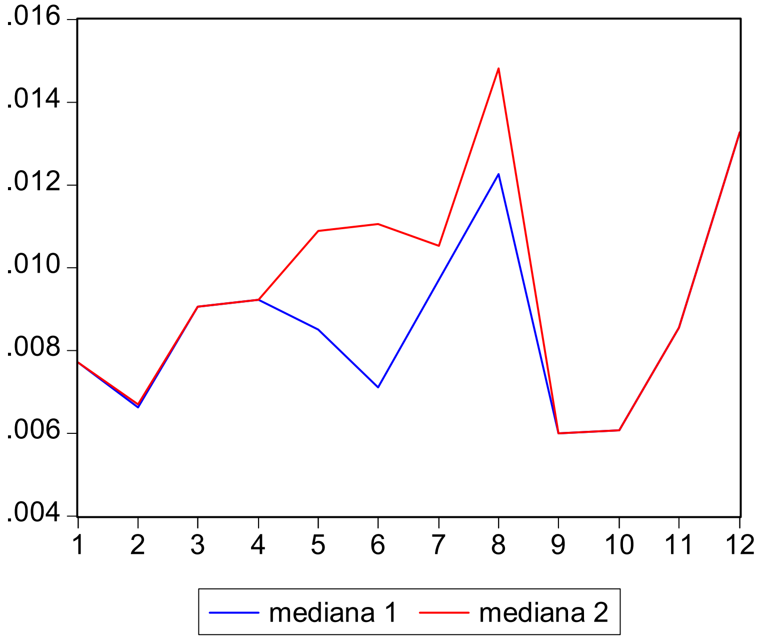


figura 23-grafico analogo a quello in figura 22 ma riguardante la mediana anziché la media

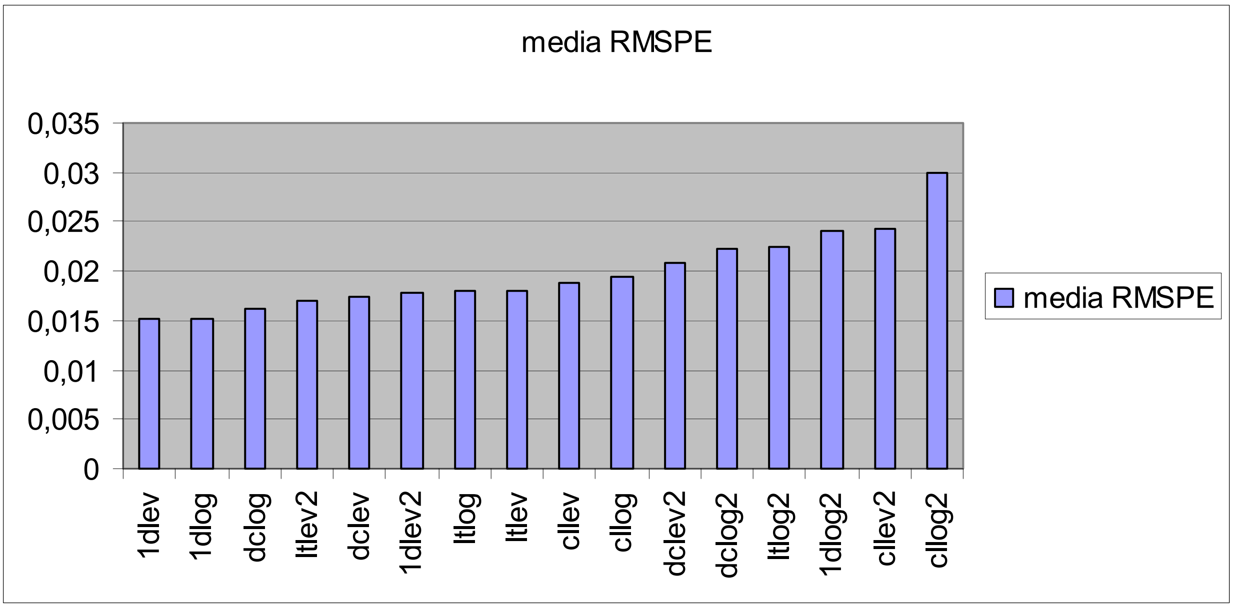
Come possiamo vedere dal grafici togliere o meno i modelli afflitti dal problema di cui si è appena parlato cambia i rendimenti, ma non ha un effetto univoco sui vari modelli, non si ha un peggioramento o un miglioramento generale, alcuni modelli risentono del cambiamento in maniera lievemente positiva e altri in maniera lievemente negativa ;la variazione è in tutti i casi comunque poco rilevante, almeno non abbastanza da giustificare la cancellazione di una parte di dati così vasta. Se si considera il grafico delle mediane degli RMSPE si può vedere come la curva rossa, ossia la mediana degli RMSPE dopo la scrematura di tutti i modelli con ϕ =0,99 e negativi tenda a stare sopra a quella blu, questo dato è piuttosto significativo e rassicura circa la possibilità di non operare questa eliminazione.

Nel paragrafo successivo vedremo comunque i risultati ottenuti in seguito a questa nuova scrematura dei dati.

5.4.2-ANALISI DEI RISULTATI

Ora analizziamo i dati ottenuti eliminando anche tutte le serie in cui il parametro AR stimato ha dato valore negativo o pari a 0,99.

I risultati in questo caso cambiano in maniera piuttosto vistosa rispetto a quelli visti in precedenza.



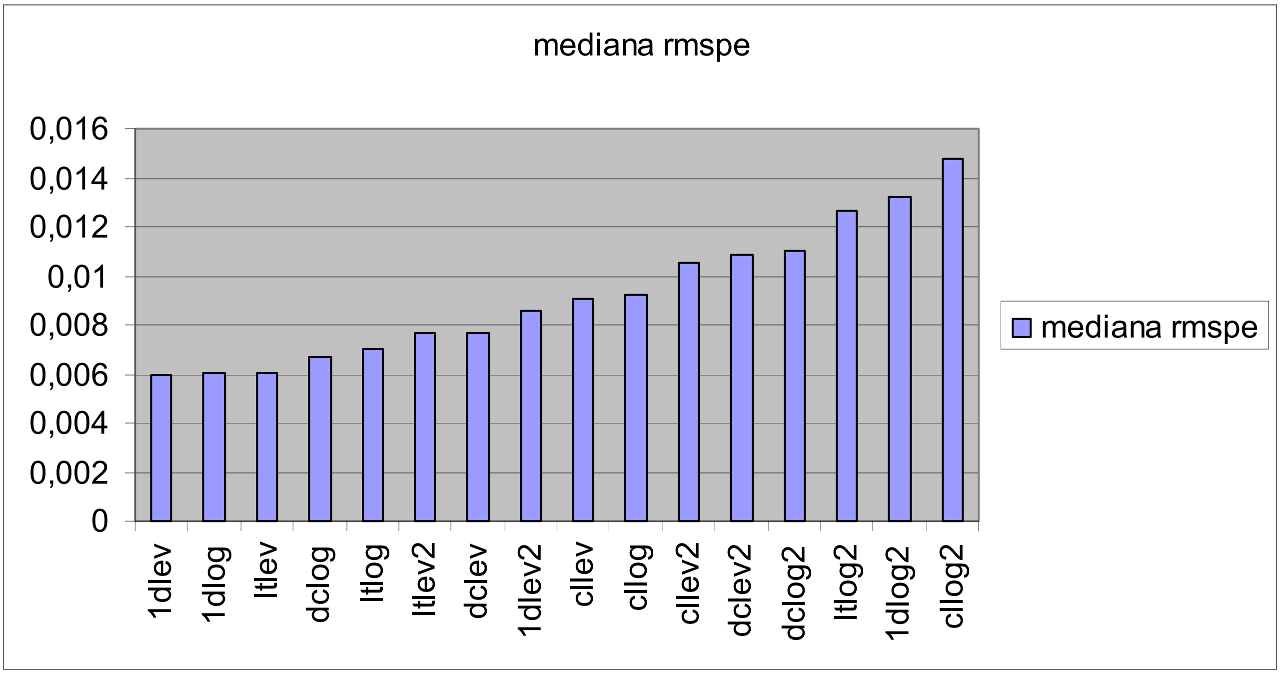
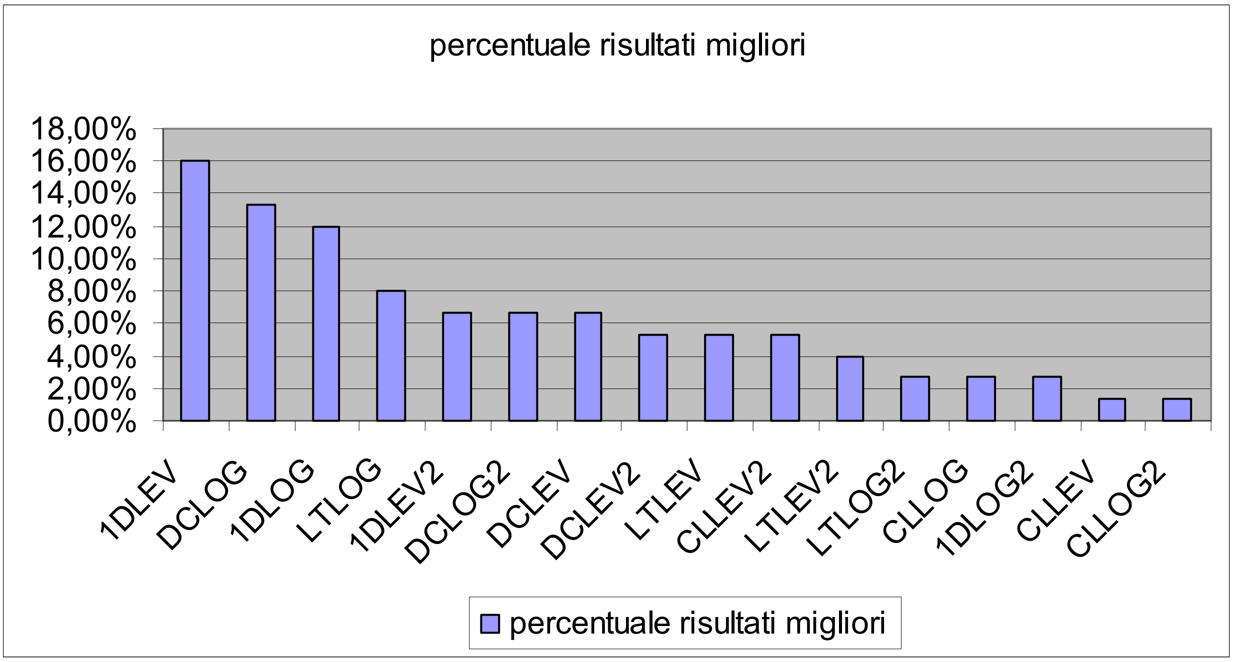


figura 24-media e mediana degli RMSPE



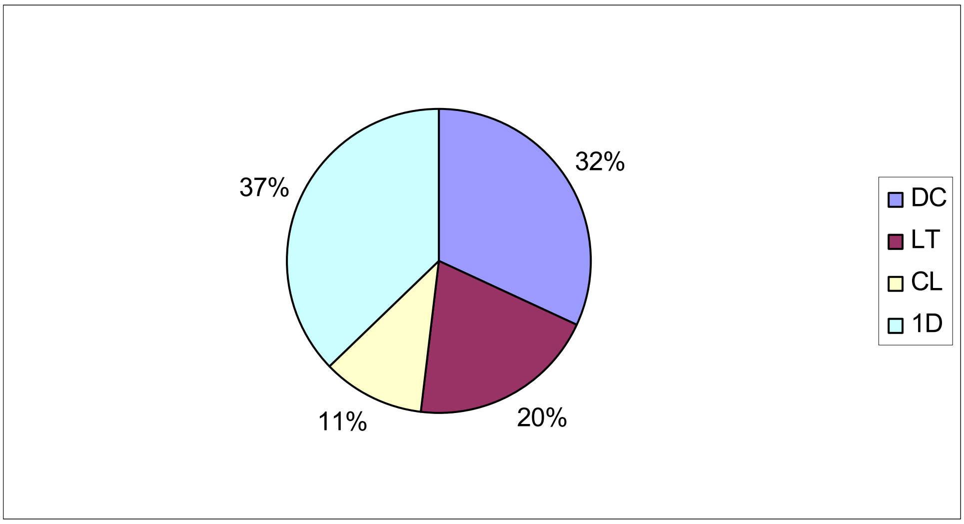


figura 25-percentuale di risultati migliori sul totale di 75 divisi per singoli modelli e per i quattro principali

Questa volta è il modello di Fernandez con costante calcolato nei livelli a comportarsi meglio di tutti: ha media e varianza dell'RMSPE minime e nel 16% dei casi ha prodotto le stime migliori. L'unico procedimento che ha avuto una performance simile è la variante logaritmica del suddetto modello.

Le stime ottenute applicando l' ADL(1,0) con costante, sia in forma logaritmica che calcolato nei livelli, forniscono ottimi risultati, pur non eguagliando le prestazioni ottenute con Fernandez.

Da notare il netto calo nelle performance delle stime ottenute coi modelli di Litterman, in particolar modo quelli con costante; prima della pesante caduta di dati avuta in questa fase di scrematura questi avevano la media e la mediana dell'RMSPE migliori; erano inoltre quelli che nel maggior numero di casi approssimavano meglio i valori reali. Ora le stime ottenute secondo Litterman in ogni variante del modello raggiungono solo il 20% delle migliori stime sul totale.

Un forte deterrente ad affrontare questa fase di pulizia dei risultati è rappresentato dalla pesante caduta di dati ma questo fa emergere particolari interessanti: c'è forte connessione fra il tipo di aggregato da stimare e la quantità di modelli che danno valore di phi=0,99 o negativo, cosa che rende di fatto inservibili alcuni metodi in relazione ad alcuni aggregati: in figura 26 possiamo vedere la percentuale di dati persi in questa seconda fase di scrematura per ognuno degli aggregati.

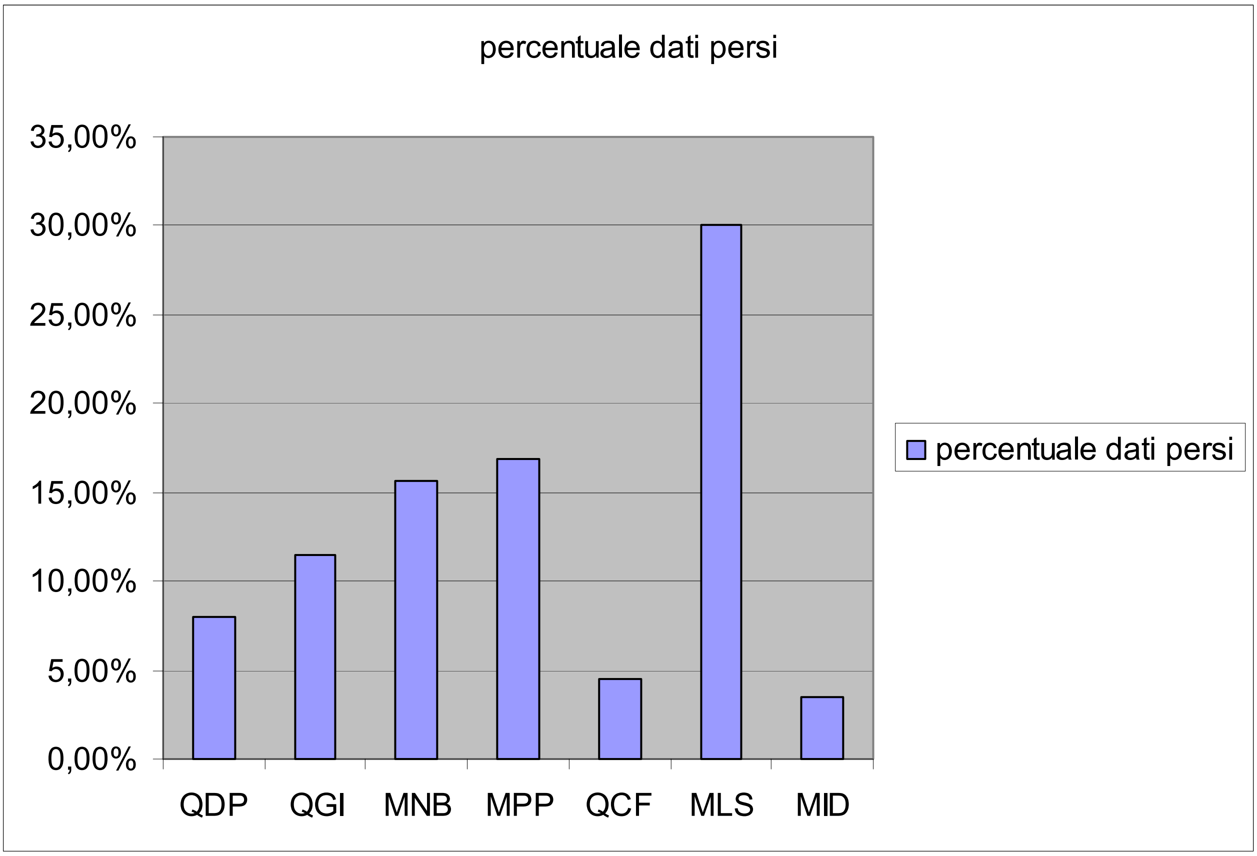


figura 26

Dalla figura 26 emerge un forte legame tra il tipo di aggregato e la quantità di dati persi, molto evidente nel caso di dell'aggregato mls che vede esattamente il 30% di dati cancellati.

Questa figura può però trarre in inganno se non si considera il pesantissimo influsso dei modelli di Litterman: in figura 27 possiamo infatti vedere la percentuale di perdita di dati dei modelli di Litterman per ognuno degli aggregati .

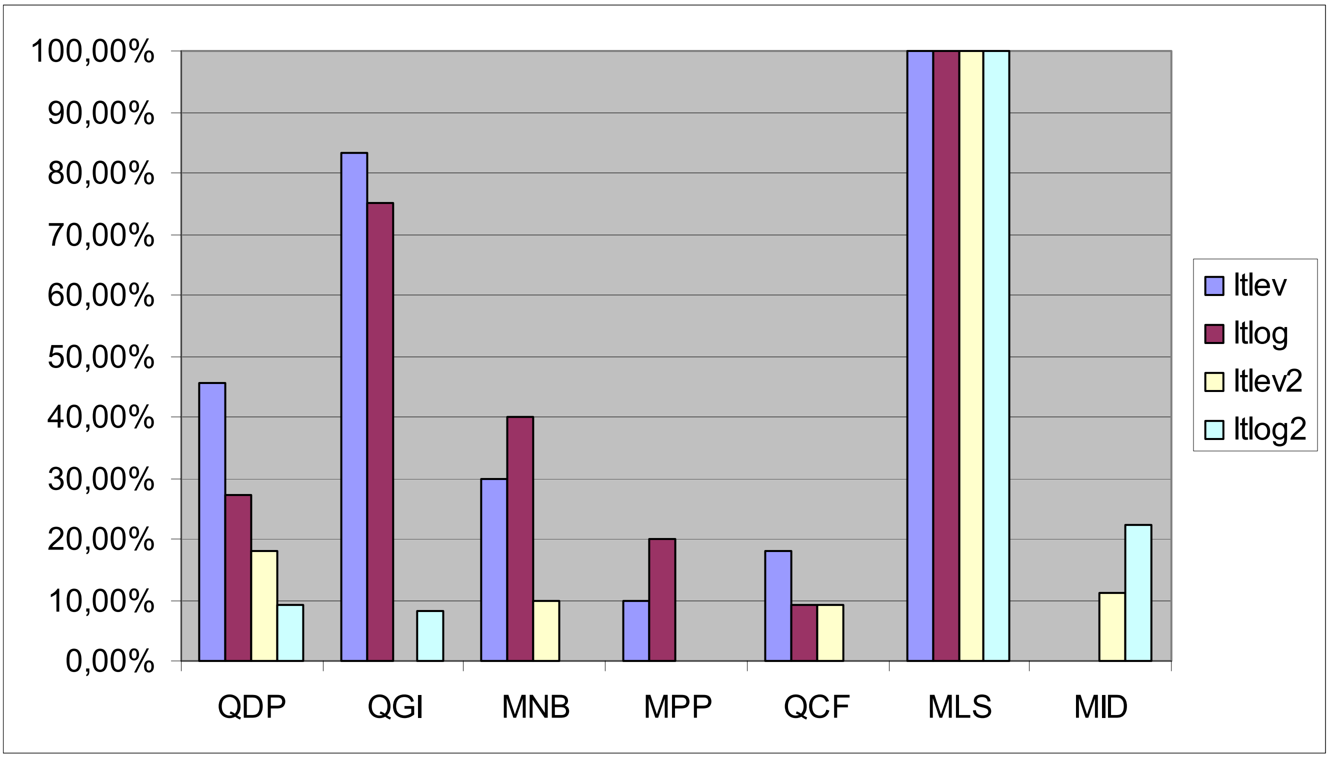


figura 27

Risulta ben chiaro che i modelli di Litterman hanno fatto registrare grosse difficoltà nella stima degli aggregati qgi e mls: per il primo dei due aggregati i modelli con costante vedono una perdita di dati prossima all'80% mentre quelli senza costante non hanno particolari problemi, ma nel secondo dei due casi la perdita di dati è pari al 100% per ognuna delle 4 tipologie di modelli.

Questo significa che ci sono serie difficoltà nell'adattamento ai dati per i modelli di Litterman e questa è una grave lacuna; una spiegazione si può trovare ma va oltre gli obbiettivi della relazione (possibili spiegazioni in Proietti 2004).

I dati più significativi di questa seconda fase di scrematura sono, come abbiamo visto, quelli che riguardano i modelli di Litterman, che si rendono quasi inutilizzabili. Gli altri modelli, almeno quelli con costante che sono quelli che sinora hanno dato risultati migliori in ognuna delle fasi di studio, non vedono particolari cambiamenti nel rendimento.

Bisogna però tenere conto che la necessità di questa seconda fase di scrematura è dubbia, dal momento che i risultati peggiorano per gran parte dei modelli come abbiamo visto al paragrafo precedente.

6-ALCUNE SERIE NEI DETTAGLI

6.1- UN CASO COMUNE

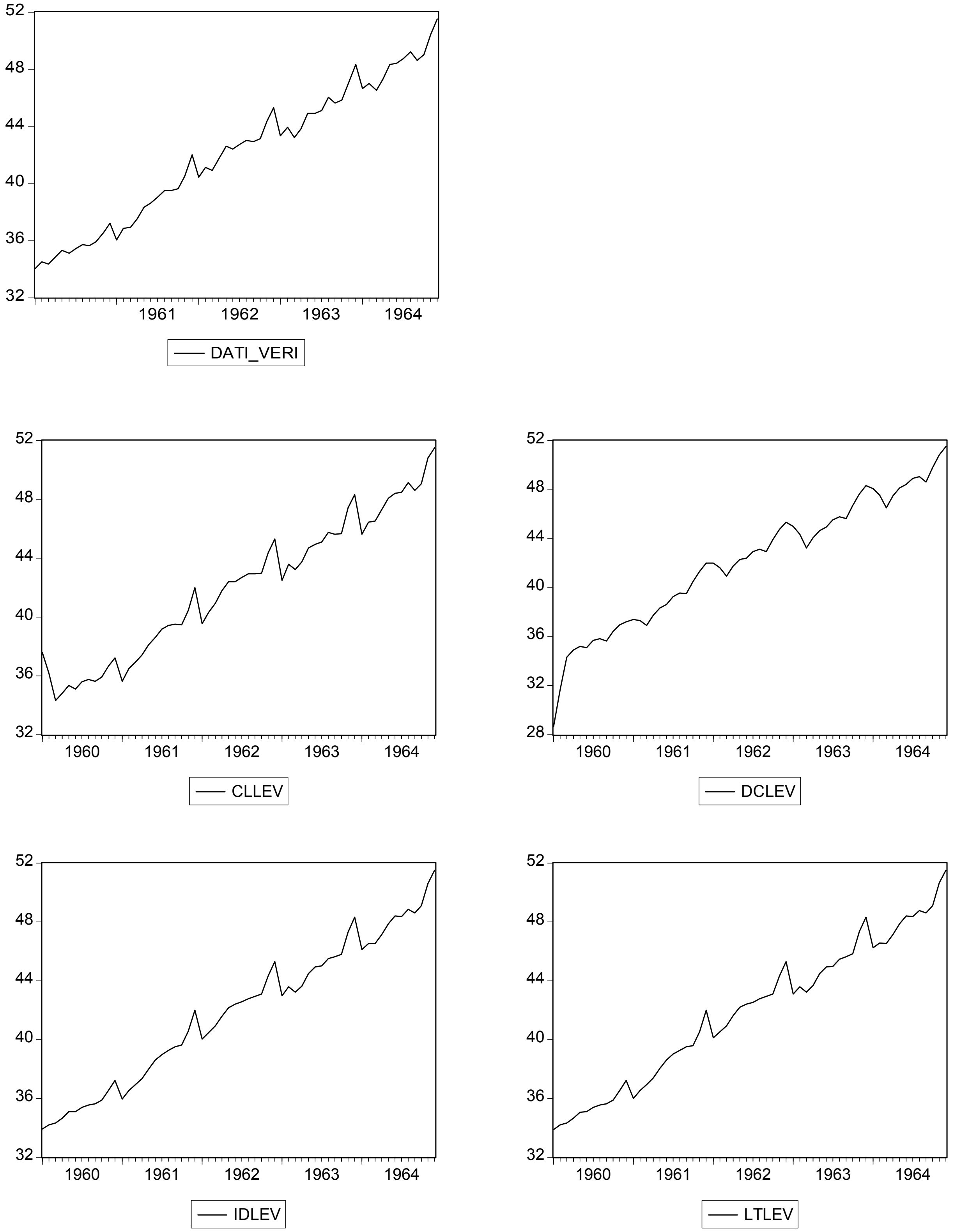


figura 1- primi cinque anni della serie Deumnb, disaggregata coi modelli con costante calcolati sui livelli, a confronto coi dati veri

In figura 1 possiamo vedere la serie Deumnb disaggregata con i 4 metodi principali, ossia tutti quelli calcolati sui livelli con costante e, più in alto, i corrispondenti valori reali. In questa e nella maggior parte delle analisi che seguiranno prenderemo in considerazione questi quattro metodi di disaggregazione, dal momento che sono risultati i più significativi negli studi sinora svolti.

La stima di questa serie vede buoni risultati per tutti i modelli anche se non eccellenti, per quelli in figura sappiamo che l'RMSPE vale rispettivamente 0,0112 , 0,0137, 0,0070 , 0,0072 . Sono valori superiori alla mediana e inferiori alla media. La graduatoria dei modelli è rispettata, nel senso che dclev e cllev hanno rendimento peggiore rispetto a 1dlev e ltlev, e questi ultimi hanno rendimenti molto simili.

La serie vede valori mensili che vanno dal primo mese del 1960 all'ultimo del 1998, ma in questi grafici il range è stato ristretto ai primi 5 anni per meglio comprendere la dinamica delle serie.

Possiamo vedere come i modelli cllev e dclev abbiano problemi di (rispettivamente) sovrastima e sottostima dei primi valori della serie, mentre gli altri due modelli danno valori più coerenti per le prime osservazioni; questa è una delle cause della differenza di RMSPE ed è una situazione che si ripropone molto spesso (vedi sez. 4.2) .

L'andamento della serie originale è grosso modo rispettato da tutti i modelli, quello che mostra differenze più evidenti è il modello dclev, che fornisce stime dalla dinamica più smussata rispetto alle altre procedure di disaggregazione: le variazioni che nella serie originale (e nelle altre stime) sono repentine, nella serie disaggregate con dclev sono evidentemente meno improvvise.

In figura 2 possiamo vedere l'errore assoluto medio per ognuno dei modelli calcolato sui saggi di variazione relativi:

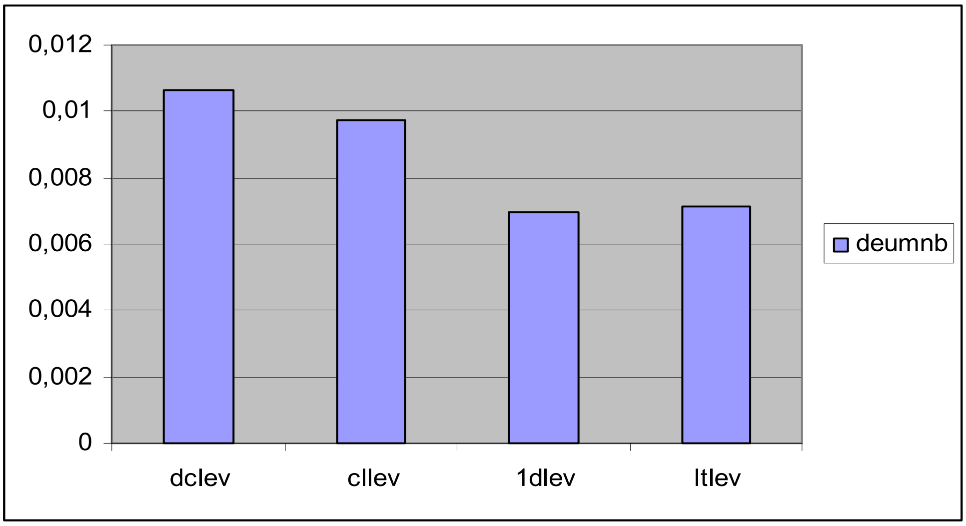


figura 2- errori assoluti medi nei saggi di variazione (relativi) riferiti ai 4 modelli principali

La graduatoria dei rendimenti per i saggi di variazione è indicativamente la stessa della graduatoria degli RMSPE, con i modelli di Litterman e Fernàndez che stimano meglio i saggi di variazione, come del resto era lecito aspettarsi data la natura dei modelli (vedi cap 3).

6.2- STIME MOLTO PRECISE

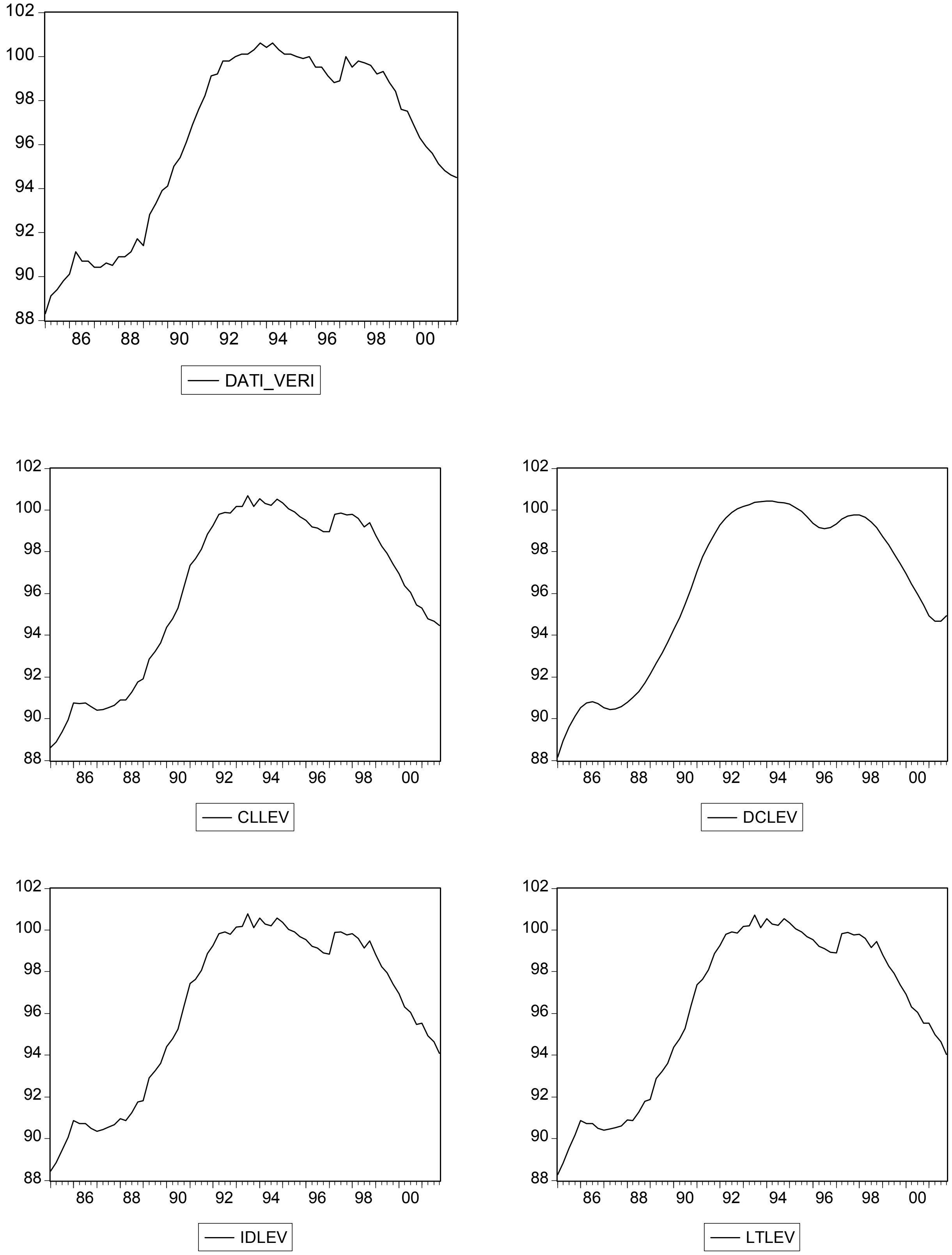


figura 3- la serie Japqdp, disaggregata coi modelli con costante calcolati sui livelli, a confronto coi dati veri

In figura 3 abbiamo la serie japqdp, una serie per cui le stime sono state ottime per tutti e 4 i modelli presi in considerazione, con RMSPE pari a (rispettivamente) 0,0021 , 0,0022 , 0,0024 , 0,0024 . Sono valori sotto la media e sotto la mediana.

Si può vedere come il grafico dei dati veri sia sostanzialmente uguale a quelli delle stime, almeno nell'andamento principale; ancora una volta nel caso di dclev le stime approssimano la serie con un andamento "ammorbidito" rispetto all'originale.Questo comportamento dei modelli dclev è molto evidente e non sembra essere un punto di forza.

In questo caso i primi due modelli hanno dato stime migliori rispetto agli altri; se la cosa non stupisce guardando il grafico delle stime fatte col metodo di Chow e Lin non si può dire altrettanto se si osserva l'andamento della serie disaggregata con dclev, dal momento che sembra essere quella che rispecchia meno i dati reali.

In figura 4 si può vedere come si comportano rispetto ai dati reali le serie disaggregate nel dettaglio in un intervallo temporale ristretto rispetto a quello in figura 2, per la precisione dal primo quadrimestre del 1992 all'ultimo del 1996 .

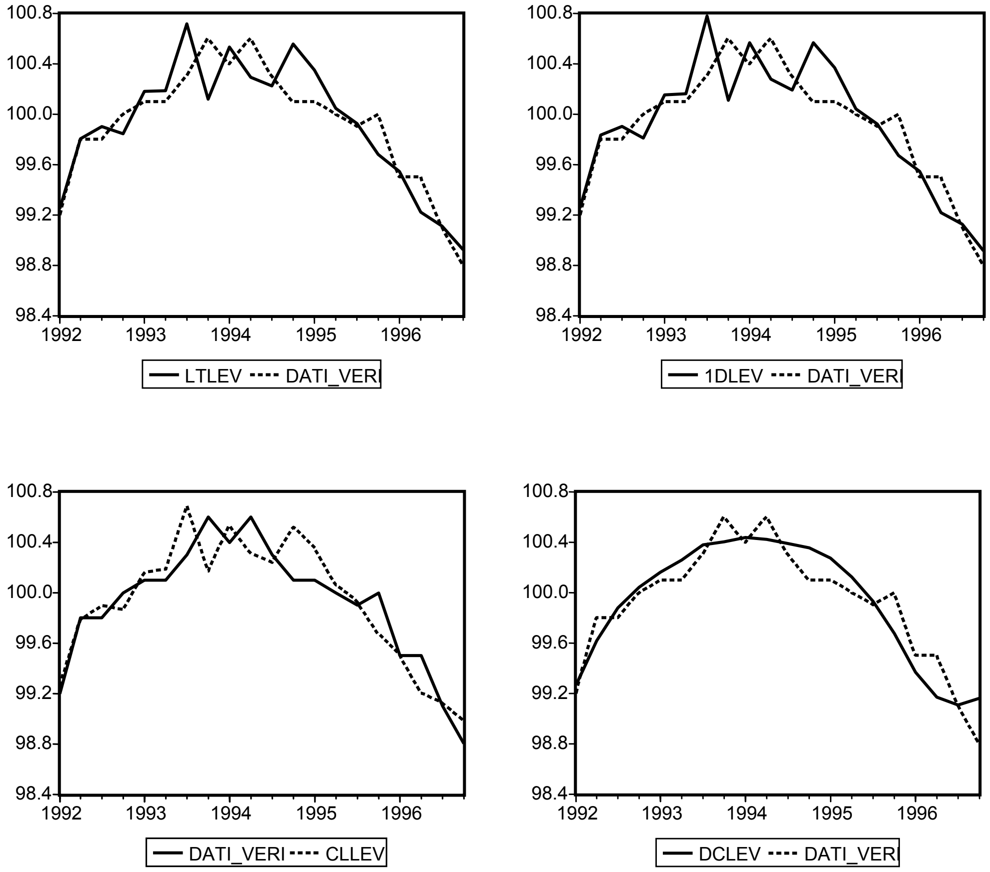


figura 4- serie japqdp disaggregata coi 4 metodi principali a confronto con i dati reali in dettaglio

Osservando i dati in figura 4 si riesce a comprendere meglio il motivo per il quale il modello dclev (almeno in questo caso) abbia un RMSPE basso : approssima l'andamento della serie reale "senza rischiare", gli altri modelli hanno infatti un andamento più simile a quello della serie, dal punto di vista visivo, ma possono produrre saggi di variazione di segno opposto a quello reale in maniera piuttosto evidente (si guardi la seconda metà del 1993 per convincersene).

Anche in questo caso per meglio comprendere come si comportano i modelli rispetto ai saggi di variazione li calcoliamo per ognuno dei metodi sopra citati (e anche per tutti gli altri)e otteniamo l'errore assoluto medio su questi indicatori ,i risultati in figura 5:

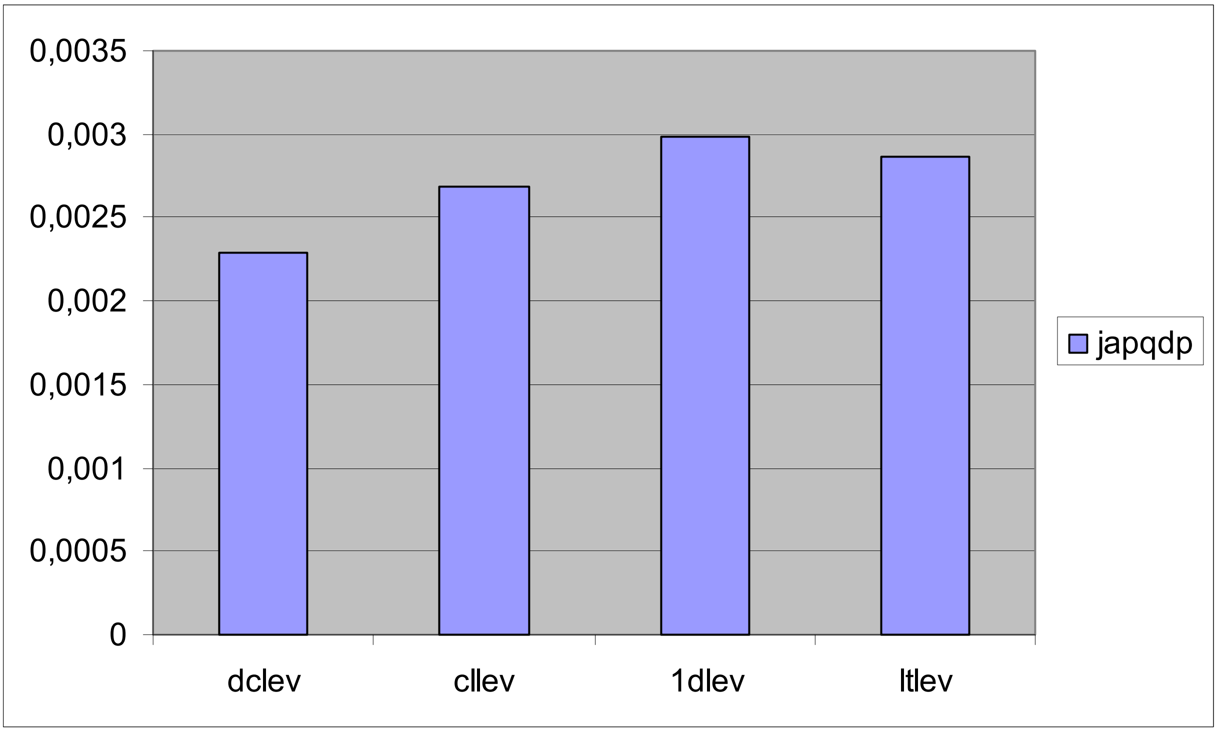


figura 5- errori assoluti medi nei saggi di variazione (relativi) riferiti ai 4 modelli principali

Curiosamente è proprio l'ADL(1,0), ossia dclev, che in questo caso ha meglio approssimato i saggi di variazione reali, mentre il modello di Fernàndez e quello di Litterman, che per costruzione teorica davano l'impressione di poter agire in maniera più efficace in questo "settore" si sono dimostrati i due più scadenti.

Per quel che riguarda i modelli non inclusi in figura 5 i risultati non differiscono molto, rimane comunque dclev il "vincitore".

6.3 – UN CASO PIU' COMPLESSO

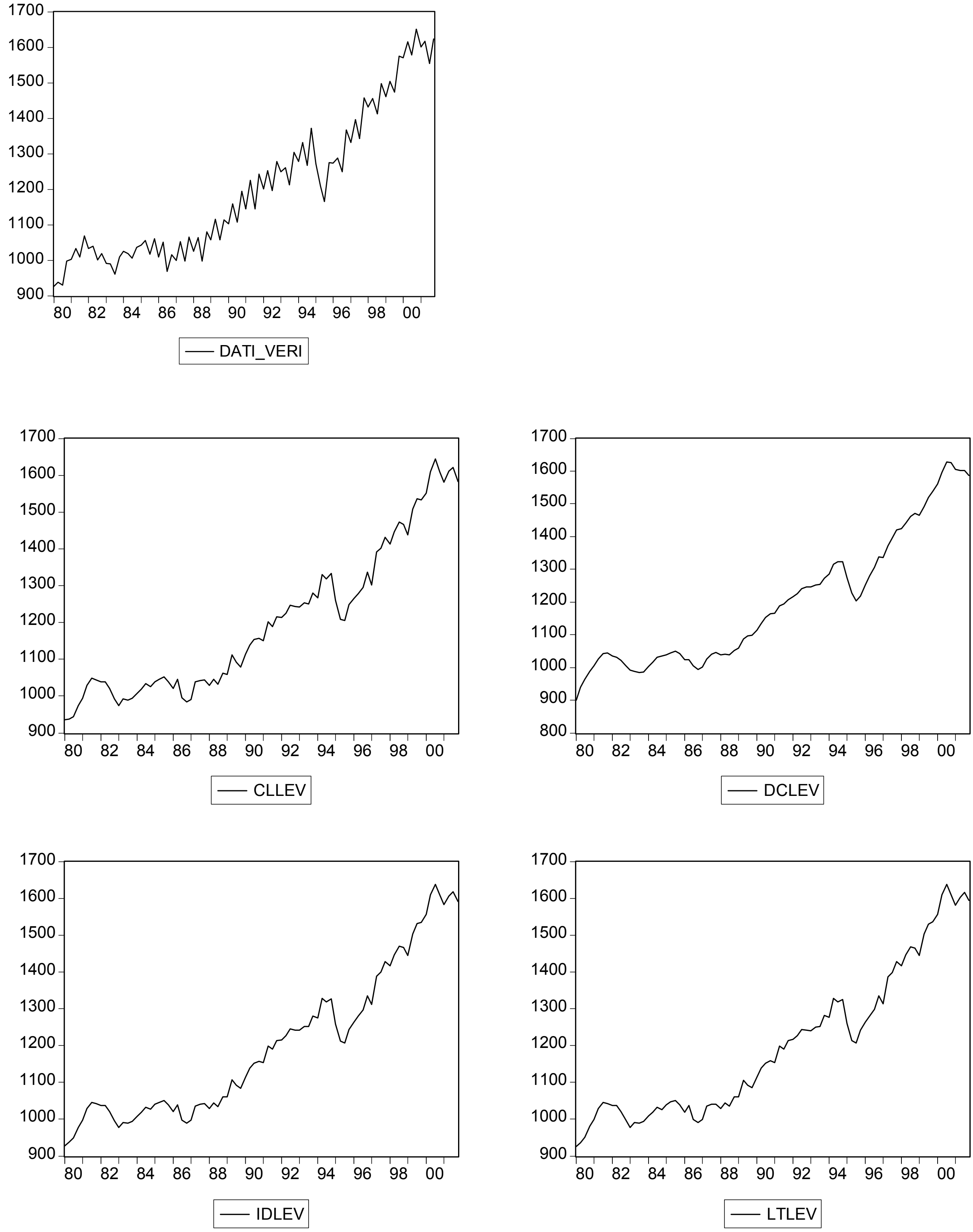


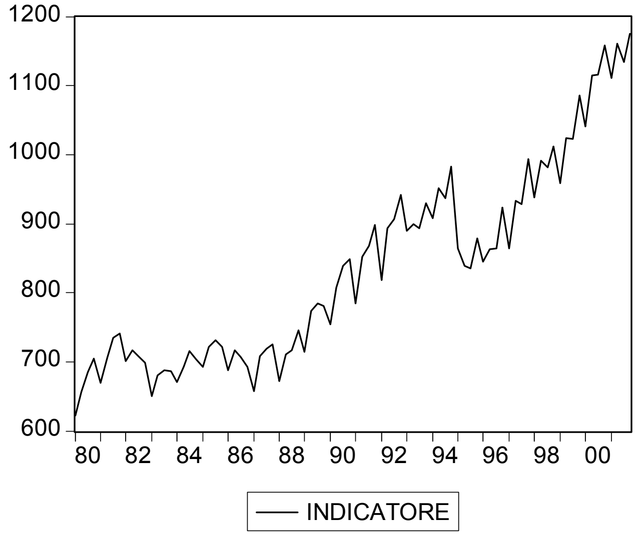
figura 6- la serie mexqgi, disaggregata coi modelli con costante calcolati sui livelli, a confronto coi dati veri

In questo caso abbiamo una situazione in cui la stima è stata meno precisa per tutti modelli.

L'RMSPE vale infatti rispettivamente 0,0228 , 0,0234 , 0,0224 , 0,0224 , ciò denota una stima all'incirca 10 volte meno precisa di quella avuta nel caso di japqdp. Sono valori ampiamente sopra la mediana e molto simili alla media.

Possiamo vedere come ancora una volta il grafico di dclev sia molto arrotondato rispetto agli altri; ma a differenza dei casi esposti in precedenza stavolta nemmeno gli altri tre modelli sono riusciti a cogliere in maniera fedele l'andamento reale della serie.

In questo caso la stima non è eccessivamente errata dal punto di vista dell'RMSPE ma nessuno dei modelli sembra aver colto la stagionalità della serie. Andiamo a vedere il grafico dell'indicatore di mexqgi in rapporto ai dati reali:



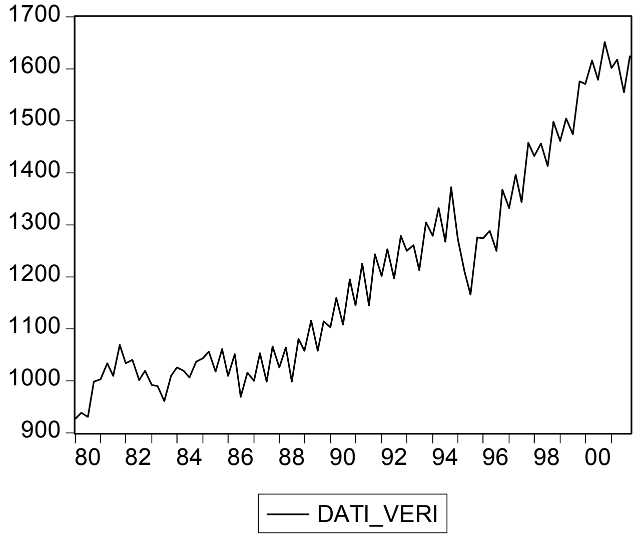


figura 7- andamento dell'indicatore di riferimento della serie mexqgi e dei dati reali

In figura 7 possiamo vedere come in effetti sia l'indicatore stesso a non cogliere appieno la stagionalità dei dati reali dal momento che le fluttuazioni di questi sembrano avere frequenza doppia rispetto a quelle dell'indicatore.

Probabilmente la causa dell'inadeguatezza delle stime è da ricercarsi nella carenza informativa dell'indicatore.

Come di consueto andiamo ad analizzare gli errori assoluti medi calcolati sui saggi di variazione:

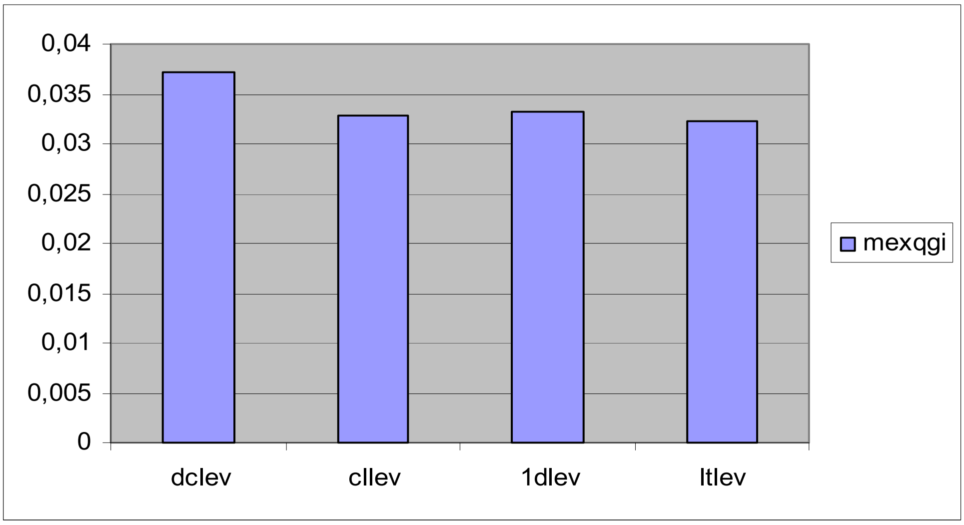


figura 8- errori assoluti medi nei saggi di variazione (relativi) riferiti ai 4 modelli principali

Per la seconda volta nei tre casi analizzati spetta a dclev il ruolo di peggior stimatore dei saggi di variazione, gli altri tre modelli hanno invece un rendimento piuttosto simile.

6.4 – STIME POCO PRECISE

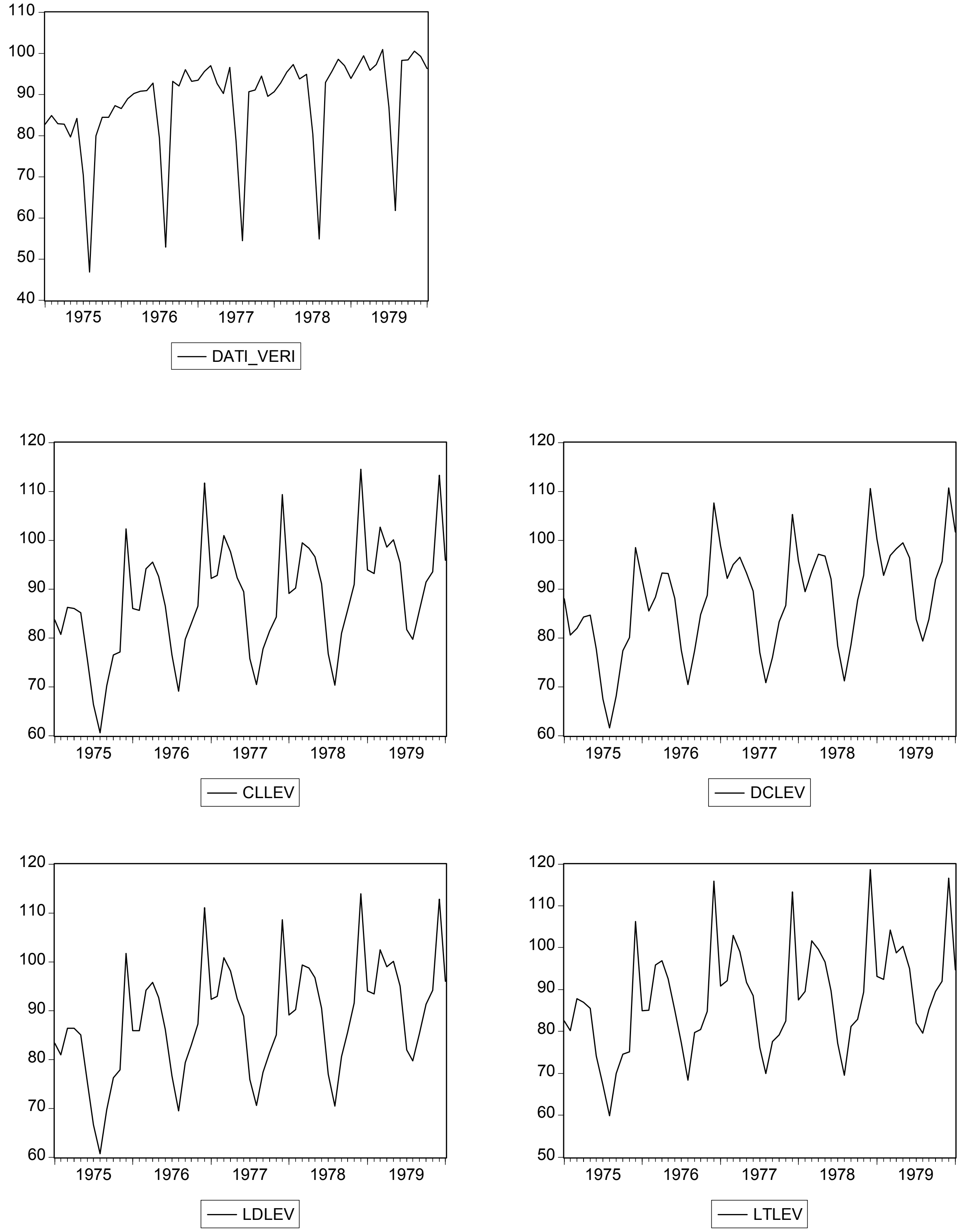


figura 9- la serie framid, disaggregata coi modelli con costante calcolati sui livelli, a confronto coi dati veri

In figura 9 viene analizzata la serie Framid nei primi 5 anni.

In questo caso gli RMSPE sono alti, rispettivamente 0,1007 , 0,0990 , 0,0999 , 0,1097 . Le stime sono le meno corrette viste sinora e lo si può vedere chiaramente in figura: c'è un calo stagionale a cavallo fra il secondo e il terzo trimestre, che nei dati reali vediamo essere più repentino e nell'ultimo trimestre c'è un forte rialzamento stagionale che nei dati reali non si realizza.

Questo perché la serie stimata è sulla produzione industriale a cadenza mensile, e questo spiega il picco discendente nei mesi estivi, ma l'indicatore di riferimento è costituito dagli ordinativi industriali, che a fine anno vedono un forte rialzo (probabilmente per le festività).

Vediamo quindi l'indicatore di riferimento :

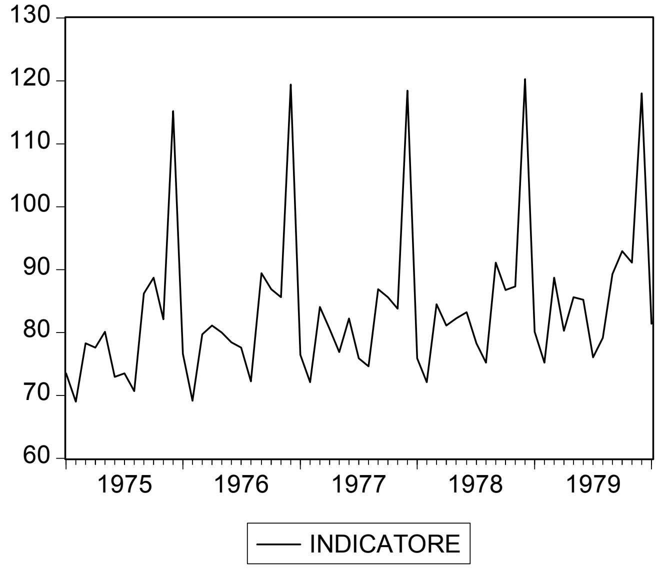


figura 10- indicatore di riferimento della serie framid

Qui il picco degli ordinativi a fine anno è piuttosto evidente e la diversità di comportamento rispetto alla serie dei dati reali lo è altrettanto; tutto questo comporta il margine di errore relativamente alto e ricorda che i metodi in questione senza un'adeguata scelta dell'indicatore di riferimento perdono (ovviamente) efficacia.

Passiamo come di consueto a vedere come si sono comportati i modelli sui saggi di variazione:

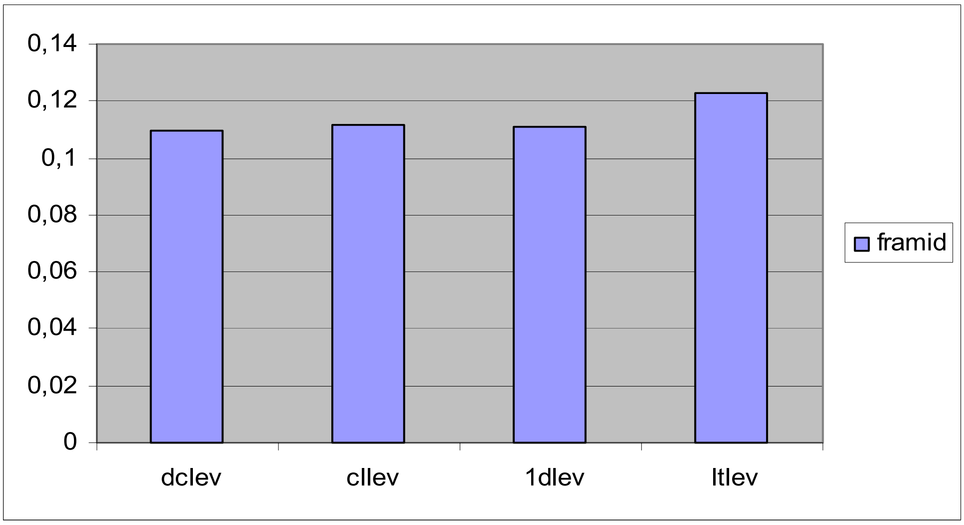


figura 11- errori assoluti medi nei saggi di variazione (relativi) riferiti ai 4 modelli principali

Dalla figura si nota subito come l'unità di misura sull'asse delle ordinate sia significativamente diversa dai grafici visti sinora, questo denota una presenza di errori nei saggi d variazione molto più consistenti in questo caso che negli altri; la considerazione risulta piuttosto ovvia se si pensa a quanto le stime siano meno precise in termini di RMSPE.

Ciò che incuriosisce è che di nuovo, come nel caso al par. 6.2, dclev è il miglior stimatore sui saggi di variazione, la cosa strana è che nei quattro casi analizzati due volte si è rivelato il migliore, viceversa le altre due volte è stato il peggiore.

Comunque non sembra esserci molta differenza fra i vari modelli e il loro rendimento sui saggi di variazione, le differenze sembrano essere imputabili alla precisione di stima in generale: i casi 6.1 e 6.2 dimostrano che sui saggi di variazione tende a riproporsi la “classifica” in termini di RMSPE, almeno in via generica.

7-VALUTAZIONI DEI MODELLI

7.1-CHOW-LIN -IL MODELLO “ISTAT” – AR(1)

In tutta l’analisi sinora svolta non si è arrivati a poter dire quale sia il modello migliore senza ombra di dubbio, perché sembra non ce ne sia uno universalmente riconoscibile come tale, ma si può affermare con certezza che il modello ufficiale utilizzato dall’ISTAT non lo è; è anzi uno di quelli con i rendimenti più scarsi. Per poterlo confrontare con la concorrenza andiamo a valutare lo scarto percentuale fra gli RMSPE medi dei modelli tenendo come punto di riferimento cllev (i valori sono calcolati nella prima fase di scrematura)

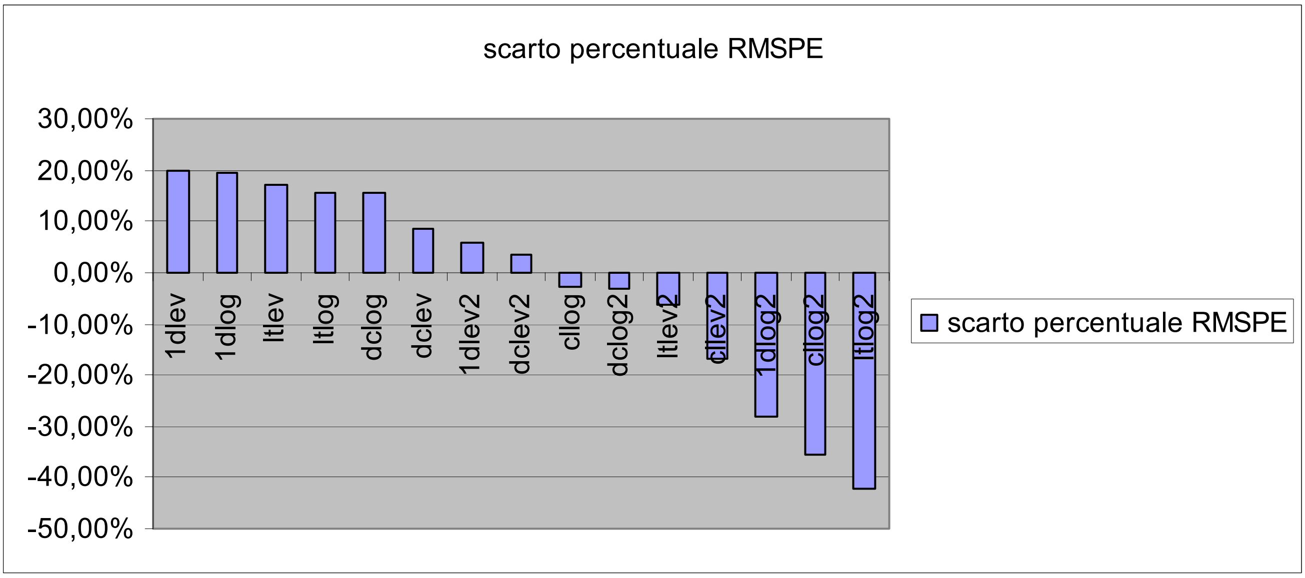


figura 1-differenza percentuale fra l’RMSPE medio del metodo “istat” e gli altri

Dalla figura 1emergono chiaramente le carenze del modello “ufficiale” dal momento che tutti i modelli il cui grafico sta sopra la linea delle ascisse si sono rivelati più precisi di Chow-Lin nelle stime. Nessun modello con costante, al di fuori di Chow-Lin logaritmico, funziona peggio del modello “ISTAT” , addirittura lo scarto percentuale in termini di RMSPE con i modelli di Litterman e con y differenziato (nei modelli presi con costante) è fra il 15 e il 20 % e solamente in una delle 75 serie disaggregate le stime di cllev sono state migliori di quelle di tutti gli altri modelli.

È inoltre evidente una certa tendenza del modello a dare stime errate nei primi valori.

Il metodo in questione ha il pregio di dare poche volte un phi pari a 0,99 , che lo rende potenzialmente più sicuro, anche se meno efficace, dei modelli di Litterman, che offrono ottime stime, ma ,se venisse rispettata la regola di eliminare per le quali la procedura di stima non converge, sarebbero inutilizzabili a causa della perdita troppo consistente di dati.

I modelli con y differenziato, in special modo quello calcolato sui livelli con costante, danno sempre stime migliori; risulta quindi piuttosto difficile giustificare l’uso del modello di Chow Lin con costante nei livelli, almeno alla luce dei risultati ottenuti sul database oggetto di studio.

Ovviamente nel nostro studio il modello di Chow Lin è stato visto in quattro forme differenti, quella appena trattata è la versione “ufficiale” nonché la migliore per quel che riguarda i risultati, le altre tre forme non mostrano caratteristiche interessanti di alcun genere.

7.2- ADL(1,0)

I modelli dc si sono piazzati sempre dignitosamente ma non si sono mai distinti in maniera particolare; sono quelli maggiormente afflitti dal problema delle “stime illogiche” date dal software, e comunque si tratta di un numero limitato di casi. Anche in questa tipologia di modelli è riscontrabile una certa tendenza all’errore in fase di stima per quel che riguarda i primi valori, e abbiamo avuto modo di notare ampiamente nel capitolo precedente come le serie disaggregate con questi metodi abbiano un comportamento “ammorbidito”. Questo punto può essere un grave problema, dal momento che in economia è molto importante comprendere la dinamica dei saggi di variazione; ma abbiamo visto come in effetti questo atteggiamento anomalo delle stime non sia stato negativo, addirittura in due dei quattro casi visti in precedenza dclev si è rivelato il miglior stimatore dei saggi di variazione.

È positiva la tendenza a non dare stime pessime in nessuna forma, solamente il modello senza costante calcolato sui logaritmi si comporta peggio del modello ISTAT , ma di poco (3,3%) ; ed è l’unico che non raggiunge il rendimento di cllev. Eppure anche questo modello, pur essendo calcolato sui logaritmi senza costante, quindi appartenendo alla tipologia meno efficace, ha una caratteristica che lo fa notare: è terzo nella classifica che mostra quante volte sul totale ha prodotto le migliori stime ma ha RMSPE medio piuttosto alto,ciò significa che alterna rendimenti ottimi ad altri, in numero maggiore, scadenti.

Il modello logaritmico con costante sembra funzionare meglio di quello calcolato sui livelli, è questo è curioso se si considera che la trasformazione logaritmica è stata fatta per ogni serie senza prendere in considerazione la reale necessità di questa.

Per quanto riguarda le serie con un phi pari a 0,99 i modelli dc, almeno quelli più efficienti, non hanno particolari problemi, quelli senza costante invece perdono una quantità piuttosto significativa di dati, intorno al 20%, ma è una caratteristica comune a tutte le tipologie di modelli calcolati senza costante.

7.3-LITTERMAN – ARIMA (1,1,0)

I modelli di Litterman sono fra i più efficaci, ma danno un discreto numero di problemi. Le stime date dai modelli con costante sono significativamente migliori di quelle dati da quelli senza costante, seguendo una tendenza di tutti i modelli ma accentuandola. Il modello ltlog2, ossia calcolato sui logaritmi senza costante ha dato risultati molto scarsi, cosa che ha influenzato il giudizio sui modelli di Litterman quando sono stati valutati nel loro insieme. Comunque, come già detto, non c'è stato nessun criterio di scelta sul come e perché applicare le trasformate logaritmiche, è stato fatto in ognuna delle serie, e questo è un procedimento poco sensato e giustifica in parte questi rendimenti scadenti. I modelli con costante hanno sempre prodotto ottime stime, anche nei casi in cui era risultato un phi pari a 0,99; questo è il punto debole maggiore dei modelli di Litterman, la difficoltà di applicazione e la difficile adattabilità ai dati: è si vero che dalla valutazione del paragrafo 5.4.1 non emerge una chiara necessità di togliere le stime prodotte in caso di phi=0,99 , ma dal punto di vista statistico sarebbe la cosa più giusta da fare, e i modelli di Litterman si trovano moltissime volte in questa situazione, in special modo quelli calcolati con la costante. Per l'aggregato MLS addirittura nessuno dei modelli di Litterman è mai riuscito a produrre un phi diverso da 0,99 , eppure proprio per l'aggregato in questione il valore minimo di RMSPe apparteneva a ltlog.

Hanno inoltre la sgradevole prerogativa di richiedere una quantità di tempo in sede di elaborazione su processore eccessiva se paragonata agli altri modelli. Questi modelli hanno dato un altro problema al computer: per ottenere le stime è stata eseguita una procedura di stima iterativa la cui precisione aumenta all'aumentare del numero di "passi", che possono arrivare ad un massimo di 501, ma per Litterman non è stato possibile settar l'opzione scan a 501 perché il programma incorreva in un errore.La stima è stata fatta quindi con un numero minore di passi.

Rimane indubbio comunque il valore del dato che ci dice che il 48% delle serie disaggregate ha avuto la migliore stima seguendo le procedure di Litterman, nel 34,6% dei casi sono stati i modelli con costante a produrre queste stime.

7.4-MODELLI CON DIFFERENZA PRIMA (FERNÀNDEZ)- random walk

I risultati ottenuti applicando il metodo esposto al 3.2 sembrano essere i migliori: quelli con costante si alternano alla “leadership” delle classifiche per media e mediana con i modelli di Litterman, prima e dopo le fasi di scrematura. Non spiccano invece i dati che riguardano la percentuale di stime migliori sul totale, ciò denota una generale tendenza di questa tipologia di modelli a non dare stime lontane dalla realtà, anche se non le migliori.

Come al solito le regressioni calcolate con la costante hanno una migliore performance, in questo caso in maniera piuttosto marcata.

Non esiste per questa categoria di modelli il problema del phi=0,99 e questa caratteristica è forse il punto forte, dal momento che in qualunque situazione di scrematura il numero di dati che rimangono è più alto rispetto a tutta la concorrenza. Oltretutto in nessun caso gli output prodotti dai software erano corrotti da valori assurdi, è quindi ridotto al minimo il rischio di consistenti riduzioni dei risultato.

Secondo le premesse teoriche (3.2) disaggregare con il metodo di Fernàndez (o con quello di Litterman) è paragonabile ad applicare una regressione sui saggi di variazione, ma nel capitolo precedente abbiamo visto che questo non implica necessariamente un miglior rendimento in termini di risultati .

7.5-CONCLUSIONI

Alla luce dei risultati ottenuti non si può affermare che uno dei metodi analizzati sia universalmente il migliore, possiamo però dare delle indicazioni che rendono più chiara la situazione.

Un occhio di riguardo era per i modelli dc, ossia l'ADL(1,0), tuttora in fase di implementazione su processore; noi ci siamo appoggiati ad un programma, DynChow, non ufficiale, comunque non in forma definitiva.

Questi metodi si sono comportati relativamente bene, hanno dimostrato significative differenze rispetto alla concorrenza e presentavano alcune peculiarità, certe positive altre meno. L'utilizzo di questi al momento appare ancora un po' prematuro, più che altro perché necessitano di un attento controllo dei dati alla ricerca delle stime che non rispettano i criteri enunciati nella sezione 5.2; e questo fatto richiede un dispendio di tempo eccessivo. Nelle versioni definitive dei programmi atti a produrre le stime il problema potrebbe essere risolto con delle procedure più efficace.

Ad ogni modo le due categorie di modelli che si sono rivelate più convincenti sono state 1d e lt, quelli di Litterman e di Fernàndez .

I risultati dei due sono molto simili, come del resto lo sono le premesse teoriche, ci sono tuttavia un paio di punti che fanno propendere verso la soluzione proposta da Fernàndez. Innanzitutto la disaggregazione coi metodi di Litterman ha richiesto un' intera notte in fase di elaborazione, con un processore a 2.4 Ghz, questo è probabilmente conseguenza delle procedure del programma ma è anche effetto dell'imponente quantità di regressioni che il computer deve svolgere.

Le regressioni di Litterman inoltre si trovano spesso danno una stima del parametro autoregressivo pari a 0,99 , dimostrando quindi che il modello ha difficoltà nello stimare i dati,o meglio che i dati si adattano a fatica al modello stesso; curiosamente però i risultati non sembrano risentirne e anche nelle serie in questione le stime prodotte sono valide.

In sostanza i due modelli si equivalgono ma quello di Fernàndez non ha dato alcun genere di problema e sembra ragionevole che la scelta del migliore ricada su di questo.

Il metodo di Chow e Lin infine, applicato in tutte le sue forme, non ha mai dato buoni risultati; ha dimostrato scarsa capacità di adattamento ai dati e spesso le stime prodotte mostrano valori vistosamente anomali. Come già detto diversi enti statistici utilizzano questo metodo e la cosa risulta strana se si considera che spesso esso ha fornito risultati peggiori rispetto agli altri metodi analizzato in questo lavoro sembra essere più affidabile. Ogni procedura però prima di poter essere effettivamente applicata su larga scala deve essere facilmente utilizzabile e al momento (penso) non è possibile ottenere dati migliori da programmi specifici in maniera rapida e facilmente comprensibile.

RIFERIMENTI BIBLIOGRAFICI

Bournay J. and Laroque G., “Réflexions sur la méthode d'élaboration des comptes trimestriels”, *Annales de l'INSEE*, 1979, *36*, pp. 3-30.

Chow G. and Lin A.L., “Best linear unbiased interpolation, distribution and extrapolation of time series by related series”, *The Review of Economics and Statistics*, 1971, *53*, pp. 372-375.

Denton F.T., “Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization”, *Journal of the American Statistical Association*, 1971, *66*, pp. 99-102.

Di Fonzo T., *La stima indiretta di serie economiche trimestrali*, Padova: Cleup, 1987.

Di Fonzo T., “The estimation of *M* disaggregate time series when contemporaneous and temporal aggregates are known”, *The Review of Economics and Statistics*, 1990, *72*, pp. 178-182.

Di Fonzo T., *Temporal disaggregation of a system of time series when the aggregate is known. Optimal vs. adjustment methods*, INSEE-Eurostat Quarterly National Accounts workshop, Paris-Bercy, December 1994, Luxembourg, Eurostat, 2002a, pp. 63-77.

Di Fonzo T., “Temporal disaggregation of economic time series: towards a dynamic extension”, Università degli Studi di Padova, Dipartimento di Scienze Statistiche, Working Paper 2002.17, 2002b.

Fernández R.B., “A methodological note on the estimation of time series”, *The Review of Economics and Statistics*, 1981, *63*, pp. 471-478.

Litterman R.B., “A random walk, Markov model for the distribution of time series”, *Journal of Business and Economic Statistics*, 1983, *1*, pp. 169-173.

Moauro G. and Savio G., “Disaggregation of time series using common components models”, mimeo, 2004.

Salazar E.L., Smith R.J. and Weale M., “Interpolation using a dynamic regression model: specification and Monte Carlo properties”, Discussion Paper n. 126, NIESR, 1997.

Santos Silva J.M.C. and Cardoso F.N., “The Chow-Lin method using dynamic models”, *Economic Modelling*, 2001, *18*, pp. 269-280.

APPENDICE

LEGENDA

Ognuno dei modelli utilizzati all'interno di questo lavoro è identificato da una stringa di caratteri del tipo xxzzz, dove:

- xx - metodo :
  1. dc = Dynamic Chow e Lin – ADL(1,0)
  2. cl = Chow e Lin – AR(1)
  3. 1d = Differenza prima – RANDOM WALK
  4. lt = Litterman – ARIMA(1,1,0)

-zzz- indica che tipo di modello all'interno di ogni metodo si sta usando :
  1. LEV = livelli con costante(o con drift nel caso 1d)
  2. LOG = logaritmi con costante (o con drift).
  3. Quando viene aggiunto un 2 alla fine vuol dire che il modello si considera senza costante (o senza drift).

Esempio: dclog2 = ADL (1,0) calcolato sui logaritmi senza costante

Le serie disaggregate sono identificate da una stringa di caratteri del tipo xxxyzz, dove:

-xxx- paese :

1- Usa- Stati Uniti d'America
2- Can- Canada
3- Aus- Australia
4- Esp- Spagna
5- Nld- Olanda
6- Gbr- Inghilterra
7- Ita- Italia
8- Deu- Germania
9- Mex- Messico
10- Fra- Francia
11- Kor- Korea
12- Jap- Giappone

-y- cadenza temporale:

1- m- mensile
2- q- trimestrale

-zzz- tipo di aggregato :

1- dp- deflatore del pil
2- gi- GDP
3- nb- broad money
4- pp- prezzi al consumo

5- cf- importazioni

6- ls- tassi interesse lungo termine

7- id- produzione industriale

8- cg- consumi

Esempio: korqdp = stime dei valori mensili del deflatore del Pil della Korea